

(19)

Europäisches Patentamt  
European Patent Office  
Office européen des brevets



(11)

**EP 0 786 519 A2**

(12)

**EUROPEAN PATENT APPLICATION**

(43) Date of publication:  
30.07.1997 Bulletin 1997/31

(51) Int Cl.<sup>6</sup>: **C12N 15/00**

(21) Application number: **97100117.7**

(22) Date of filing: **07.01.1997**

(84) Designated Contracting States:  
**AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC  
NL PT SE**  
Designated Extension States:  
**AL LT LV RO SI**

(30) Priority: **05.01.1996 US 9861**

(71) Applicant: **HUMAN GENOME SCIENCES, INC.**  
**Rockville, MD 20850-3338 (US)**

(72) Inventors:  
• **Kunsch, Charles A.**  
**Gaithersburg, Maryland 20882 (US)**

- **Chol, Gil H.**  
**Rockville, Maryland 20850 (US)**
- **Barash, Steven C.**  
**Rockville, Maryland 20850 (US)**
- **Dillon, Patrick J.**  
**Gaithersburg, Maryland 20879 (US)**
- **Fannon, Michael R.**  
**Silver Spring, Maryland 20906 (US)**
- **Rosen, Craig A.**  
**Laytonsville, Maryland 20882 (US)**

(74) Representative: **VOSSIUS & PARTNER**  
**Postfach 86 07 67**  
**81634 München (DE)**

(54) **Staphylococcus aureus polynucleotides and sequences**

(57) The present invention provides polynucleotide sequences of the genome of *Staphylococcus aureus*, polypeptide sequences encoded by the polynucleotide sequences, corresponding polynucleotides and polypeptides, vectors and hosts comprising the polynu-

cleotides, and assays and other uses thereof. The present invention further provides polynucleotide and polypeptide sequence information stored on computer readable media, and computer-based systems and methods which facilitate its use.

**EP 0 786 519 A2**

## Description

The present invention relates to the field of molecular biology. In particular, it relates to, among other things, nucleotide sequences of *Staphylococcus aureus*, contigs, ORFs, fragments, probes, primers and related polynucleotides thereof, peptides and polypeptides encoded by the sequences, and uses of the polynucleotides and sequences thereof, such as in fermentation, polypeptide production, assays and pharmaceutical development, among others.

The genus *Staphylococcus* includes at least 20 distinct species. (For a review see Novick, R. P., *The Staphylococcus* as a Molecular Genetic System, Chapter 1, pgs. 1-37 in *MOLECULAR BIOLOGY OF THE STAPHYLOCOCCI*, R. Novick, Ed., VCH Publishers, New York (1990)). Species differ from one another by 80% or more, by hybridization kinetics, whereas strains within a species are at least 90% identical by the same measure.

The species *Staphylococcus aureus*, a gram-positive, facultatively aerobic, clump-forming cocci, is among the most important etiological agents of bacterial infection in humans, as discussed briefly below.

## Human Health and *S. Aureus*

*Staphylococcus aureus* is a ubiquitous pathogen. (See, for instance, Mims *et al.*, *MEDICAL MICROBIOLOGY*, Mosby-Year Book Europe Limited, London, UK (1993)). It is an etiological agent of a variety of conditions, ranging in severity from mild to fatal. A few of the more common conditions caused by *S. aureus* infection are burns, cellulitis, eyelid infections, food poisoning, joint infections, neonatal conjunctivitis, osteomyelitis, skin infections, surgical wound infection, scalded skin syndrome and toxic shock syndrome, some of which are described further below.

### Burns

Burn wounds generally are sterile initially. However, they generally compromise physical and immune barriers to infection, cause loss of fluid and electrolytes and result in local or general physiological dysfunction. After cooling, contact with viable bacteria results in mixed colonization at the injury site. Infection may be restricted to the non-viable debris on the burn surface ("eschar"), it may progress into full skin infection and invade viable tissue below the eschar and it may reach below the skin, enter the lymphatic and blood circulation and develop into septicaemia. *S. aureus* is among the most important pathogens typically found in burn wound infections. It can destroy granulation tissue and produce severe septicaemia.

### Cellulitis

Cellulitis, an acute infection of the skin that expands from a typically superficial origin to spread below the cutaneous layer, most commonly is caused by *S. aureus* in conjunction with *S. pyogenes*. Cellulitis can lead to systemic infection. In fact, cellulitis can be one aspect of synergistic bacterial gangrene. This condition typically is caused by a mixture of *S. aureus* and microaerophilic streptococci. It causes necrosis and treatment is limited to excision of the necrotic tissue. The condition often is fatal.

### Eyelid infections

*S. aureus* is the cause of styes and of sticky eye" in neonates, among other eye infections. Typically such infections are limited to the surface of the eye, and may occasionally penetrate the surface with more severe consequences.

### Food poisoning

Some strains of *S. aureus* produce one or more of five serologically distinct, heat and acid stable enterotoxins that are not destroyed by digestive process of the stomach and small intestine (enterotoxins A-E). Ingestion of the toxin, in sufficient quantities, typically results in severe vomiting, but not diarrhoea. The effect does not require viable bacteria. Although the toxins are known, their mechanism of action is not understood.

### Joint infections

*S. aureus* infects bone joints causing diseases such as osteomyelitis.

### Osteomyelitis

*S. aureus* is the most common causative agent of haematogenous osteomyelitis. The disease tends to occur in

children and adolescents more than adults and it is associated with non-penetrating injuries to bones. Infection typically occurs in the long end of growing bone, hence its occurrence in physically immature populations. Most often, infection is localized in the vicinity of sprouting capillary loops adjacent to epiphysial growth plates in the end of long, growing bones.

#### Skin infections

*S. aureus* is the most common pathogen of such minor skin infections as abscesses and boils. Such infections often are resolved by normal host response mechanisms, but they also can develop into severe internal infections. Recurrent infections of the nasal passages plague nasal carriers of *S. aureus*.

#### Surgical Wound Infections

Surgical wounds often penetrate far into the body. Infection of such wound thus poses a grave risk to the patient. *S. aureus* is the most important causative agent of infections in surgical wounds. *S. aureus* is unusually adept at invading surgical wounds; sutured wounds can be infected by far fewer *S. aureus* cells than are necessary to cause infection in normal skin. Invasion of surgical wound can lead to severe *S. aureus* septicaemia. Invasion of the blood stream by *S. aureus* can lead to seeding and infection of internal organs, particularly heart valves and bone, causing systemic diseases, such as endocarditis and osteomyelitis.

#### Scalded Skin Syndrome

*S. aureus* is responsible for "scalded skin syndrome" (also called toxic epidermal necrosis, Ritter's disease and Lyell's disease). This disease occurs in older children, typically in outbreaks caused by flowering of *S. aureus* strains produce exfoliation (also called scalded skin syndrome toxin). Although the bacteria initially may infect only a minor lesion, the toxin destroys intercellular connections, spreads epidermal layers and allows the infection to penetrate the outer layer of the skin, producing the desquamation that typifies the disease. Shedding of the outer layer of skin generally reveals normal skin below, but fluid lost in the process can produce severe injury in young children if it is not treated properly.

#### Toxic Shock Syndrome

Toxic shock syndrome is caused by strains of *S. aureus* that produce the so-called toxic shock syndrome toxin. The disease can be caused by *S. aureus* infection at any site, but it is too often erroneously viewed exclusively as a disease solely of women who use tampons. The disease involves toxemia and septicaemia, and can be fatal.

#### Nocosomal Infections

In the 1984 National Nosocomial Infection Surveillance Study ("NNIS") *S. aureus* was the most prevalent agent of surgical wound infections in many hospital services, including medicine, surgery, obstetrics, pediatrics and newborns.

#### Resistance to drugs of *S. aureus* strains

Prior to the introduction of penicillin the prognosis for patients seriously infected with *S. aureus* was unfavorable. Following the introduction of penicillin in the early 1940s even the worst *S. aureus* infections generally could be treated successfully. The emergence of penicillin-resistant strains of *S. aureus* did not take long, however. Most strains of *S. aureus* encountered in hospital infections today do not respond to penicillin; although, fortunately, this is not the case for *S. aureus* encountered in community infections.

It is well known now that penicillin-resistant strains of *S. aureus* produce a lactamase which converts penicillin to penicilloic acid, and thereby destroys antibiotic activity. Furthermore, the lactamase gene often is propagated episomally, typically on a plasmid, and often is only one of several genes on an episomal element that, together, confer multidrug resistance.

Methicillins, introduced in the 1960s, largely overcame the problem of penicillin resistance in *S. aureus*. These compounds conserve the portions of penicillin responsible for antibiotic activity and modify or alter other portions that make penicillin a good substrate for inactivating lactamases. However, methicillin resistance has emerged in *S. aureus*, along with resistance to many other antibiotics effective against this organism, including aminoglycosides, tetracycline, chloramphenicol, macrolides and lincosamides. In fact, methicillin-resistant strains of *S. aureus* generally are multiply drug resistant.

The molecular genetics of most types of drug resistance in *S. aureus* has been elucidated (See Lyon *et al.*, *Microbiology Reviews* 51: 88-134 (1987)). Generally, resistance is mediated by plasmids, as noted above regarding penicillin resistance; however, several stable forms of drug resistance have been observed that apparently involve integration of a resistance element into the *S. aureus* genome itself.

Thus far each new antibiotic gives rise to resistance strains, strains emerge that are resistance to multiple drugs and increasingly persistent forms of resistance begin to emerge. Drug resistance of *S. aureus* infections already poses significant treatment difficulties, which are likely to get much worse unless new therapeutic agents are developed.

#### Molecular Genetics of *Staphylococcus Aureus*

Despite its importance in, among other things, human disease, relatively little is known about the genome of this organism.

Most genetic studies of *S. aureus* have been carried out using the strain NCTC8325, which contains prophages psi11 psi12 and psi13, and the UV-cured derivative of this strain, 8325-4 (also referred to as RN450), which is free of the prophages.

These studies revealed that the *S. aureus* genome, like that of other *staphylococci*, consists of one circular, covalently closed, double-stranded DNA and a collection of so-called variable accessory genetic elements, such as prophages, plasmids, transposons and the like.

Physical characterization of the genome has not been carried out in any detail. Pattee *et al.* published a low resolution and incomplete genetic and physical map of the chromosome of *S. aureus* strain NCTC 8325. (Pattee *et al.* Genetic and Physical Mapping of Chromosome of *Staphylococcus aureus* NCTC 8325, Chapter 11, pgs. 163-169 in. MOLECULAR BIOLOGY OF THE STAPHYLOCOCCI, R.P. Novick, Ed., VCH Publishers, New York, (1990) The genetic map largely was produced by mapping insertions of Tn551 and Tn4001, which, respectively, confer erythromycin and gentamicin resistance, and by analysis of SmaI-digested DNA by Pulsed Field Gel Electrophoresis ("PFGE").

The map was of low resolution; even estimating the physical size of the genome was difficult, according to the investigators. The size of the largest SmaI chromosome fragment, for instance, was too large for accurate sizing by PFGE. To estimate its size, additional restriction sites had to be introduced into the chromosome using a transposon containing a SmaI recognition sequence.

In sum, most physical characteristics and almost all of the genes of *Staphylococcus aureus* are unknown. Among the few genes that have been identified, most have not been physically mapped or characterized in detail. Only a very few genes of this organism have been sequenced. (See, for instance Thornsberry, J., *Antimicrobial Chemotherapy* 21 Suppl C: 9-16 (1988), current versions of GENBANK and other nucleic acid databases, and references that relate to the genome of *S. aureus* such as those set out elsewhere herein.)

It is clear that the etiology of diseases mediated or exacerbated by *S. aureus* infection involves the programmed expression of *S. aureus* genes, and that characterizing the genes and their patterns of expression would add dramatically to our understanding of the organism and its host interactions. Knowledge of *S. aureus* genes and genomic organization would dramatically improve understanding of disease etiology and lead to improved and new ways of preventing, ameliorating, arresting and reversing diseases. Moreover, characterized genes and genomic fragments of *S. aureus* would provide reagents for, among other things, detecting, characterizing and controlling *S. aureus* infections. There is a need therefore to characterize the genome of *S. aureus* and for polynucleotides and sequences of this organism.

The present invention is based on the sequencing of fragments of the *Staphylococcus aureus* genome. The primary nucleotide sequences which were generated are provided in SEQ ID NOS: 1-5,191.

The present invention provides the nucleotide sequence of several thousand contigs of the *Staphylococcus aureus* genome, which are listed in tables below and set out in the Sequence Listing submitted herewith, and representative fragments thereof, in a form which can be readily used, analyzed, and interpreted by a skilled artisan. In one embodiment, the present invention is provided as contiguous strings of primary sequence information corresponding to the nucleotide sequences depicted in SEQ ID NOS:1-5,191.

The present invention further provides nucleotide sequences which are at least 95%, preferably 99% and most preferably 99.9%, identical to the nucleotide sequences of SEQ ID NOS:1-5,191.

The nucleotide sequence of SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence which is at least 95%, preferably 99% and most preferably 99.9%, identical to the nucleotide sequence of SEQ ID NOS:1-5,191 may be provided in a variety of mediums to facilitate its use. In one application of this embodiment, the sequences of the present invention are recorded on computer readable media. Such media includes, but is not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media.

The present invention further provides systems, particularly computer-based systems which contain the sequence

information herein described stored in a data storage means. Such systems are designed to identify commercially important fragments of the *Staphylococcus aureus* genome.

Another embodiment of the present invention is directed to fragments, preferably isolated fragments, of the *Staphylococcus aureus* genome having particular structural or functional attributes. Such fragments of the *Staphylococcus aureus* genome of the present invention include, but are not limited to, fragments which encode peptides, hereinafter referred to as open reading frames or ORFs,\* fragments which modulate the expression of an operably linked ORF, hereinafter referred to as expression modulating fragments or EMFs,\* and fragments which can be used to diagnose the presence of *Staphylococcus aureus* in a sample, hereinafter referred to as diagnostic fragments or "DFs."

Each of the ORFs in fragments of the *Staphylococcus aureus* genome disclosed in Tables 1-3, and the EMFs found 5' to the ORFs, can be used in numerous ways as polynucleotide reagents. For instance, the sequences can be used as diagnostic probes or amplification primers for detecting or determining the presence of a specific microbe in a sample, to selectively control gene expression in a host and in the production of polypeptides, such as polypeptides encoded by ORFs of the present invention, particular those polypeptides that have a pharmacological activity.

The present invention further includes recombinant constructs comprising one or more fragments of the *Staphylococcus aureus* genome of the present invention. The recombinant constructs of the present invention comprise vectors, such as a plasmid or viral vector, into which a fragment of the *Staphylococcus aureus* has been inserted.

The present invention further provides host cells containing any of the isolated fragments of the *Staphylococcus aureus* genome of the present invention. The host cells can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic cell, such as a yeast cell, or a procaryotic cell such as a bacterial cell.

The present invention is further directed to polypeptides and proteins, preferably isolated polypeptides and proteins, encoded by ORFs of the present invention. A variety of methods, well known to those of skill in the art, routinely may be utilized to obtain any of the polypeptides and proteins of the present invention. For instance, polypeptides and proteins of the present invention having relatively short, simple amino acid sequences readily can be synthesized using commercially available automated peptide synthesizers. Polypeptides and proteins of the present invention also may be purified from bacterial cells which naturally produce the protein. Yet another alternative is to purify polypeptide and proteins of the present invention can from cells which have been altered to express them.

The invention further provides polypeptides, preferably isolated polypeptides, comprising *Staphylococcus aureus* epitopes and vaccine compositions comprising such polypeptides. Also provided are methods for vaccinating an individual against *Staphylococcus aureus* infection.

The invention further provides methods of obtaining homologs of the fragments of the *Staphylococcus aureus* genome of the present invention and homologs of the proteins encoded by the ORFs of the present invention. Specifically, by using the nucleotide and amino acid sequences disclosed herein as a probe or as primers, and techniques such as PCR cloning and colony/plaque hybridization, one skilled in the art can obtain homologs.

The invention further provides antibodies which selectively bind polypeptides and proteins of the present invention. Such antibodies include both monoclonal and polyclonal antibodies.

The invention further provides hybridomas which produce the above-described antibodies. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody.

The present invention further provides methods of identifying test samples derived from cells which express one of the ORFs of the present invention, or a homolog thereof. Such methods comprise incubating a test sample with one or more of the antibodies of the present invention, or one or more of the DFs or antigens of the present invention, under conditions which allow a skilled artisan to determine if the sample contains the ORF or product produced therefrom.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the above-described assays.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the antibodies, antigens, or one of the DFs of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of bound antibodies, antigens or hybridized DFs.

Using the isolated proteins of the present invention, the present invention further provides methods of obtaining and identifying agents capable of binding to a polypeptide or protein encoded by one of the ORFs of the present invention. Specifically, such agents include, as further described below, antibodies, peptides, carbohydrates, pharmaceutical agents and the like. Such methods comprise steps of: (a) contacting an agent with an isolated protein encoded by one of the ORFs of the present invention; and (b) determining whether the agent binds to said protein.

The present genomic sequences of *Staphylococcus aureus* will be of great value to all laboratories working with this organism and for a variety of commercial purposes. Many fragments of the *Staphylococcus aureus* genome will be immediately identified by similarity searches against GenBank or protein databases and will be of immediate value to *Staphylococcus aureus* researchers and for immediate commercial value for the production of proteins or to control gene expression.

The methodology and technology for elucidating extensive genomic sequences of bacterial and other genomes

has and will greatly enhance the ability to analyze and understand chromosomal organization. In particular, sequenced contigs and genomes will provide the models for developing tools for the analysis of chromosome structure and function, including the ability to identify genes within large segments of genomic DNA, the structure, position, and spacing of regulatory elements, the identification of genes with potential industrial applications, and the ability to do comparative genomic and molecular phylogeny.

FIGURE 1 is a block diagram of a computer system (102) that can be used to implement computer-based systems of present invention.

FIGURE 2 is a schematic diagram depicting the data flow and computer programs used to collect, assemble, edit and annotate the contigs of the *Staphylococcus aureus* genome of the present invention. Both Macintosh and Unix platforms are used to handle the AB 373 and 377 sequence data files, largely as described in Kerlavage *et al.*, *Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System Sciences*, 585, IEEE Computer Society Press, Washington D.C. (1993). Factura (AB) is a Macintosh program designed for automatic vector sequence removal and end-trimming of sequence files. The program Loadis runs on a Macintosh platform and parses the feature data extracted from the sequence files by Factura to the Unix based *Staphylococcus aureus* relational database. Assembly of contigs (and whole genome sequences) is accomplished by retrieving a specific set of sequence files and their associated features using extrseq, a Unix utility for retrieving sequences from an SQL database. The resulting sequence file is processed by seq\_filter to trim portions of the sequences with more than 2% ambiguous nucleotides. The sequence files were assembled using TIGR Assembler, an assembly engine designed at The Institute for Genomic Research (TIGR) for rapid and accurate assembly of thousands of sequence fragments. The collection of contigs generated by the assembly step is loaded into the database with the lassie program. Identification of open reading frames (ORFs) is accomplished by processing contigs with zorf. The ORFs are searched against *S. aureus* sequences from Genbank and against all protein sequences using the BLASTN and BLASTP programs, described in Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990). Results of the ORF determination and similarity searching steps were loaded into the database. As described below, some results of the determination and the searches are set out in Tables 1-3.

The present invention is based on the sequencing of fragments of the *Staphylococcus aureus* genome and analysis of the sequences. The primary nucleotide sequences generated by sequencing the fragments are provided in SEQ ID NOS:1-5,191. (As used herein, the "primary sequence" refers to the nucleotide sequence represented by the IUPAC nomenclature system.)

In addition to the aforementioned *Staphylococcus aureus* polynucleotide and polynucleotide sequences, the present invention provides the nucleotide sequences of SEQ ID NOS:1-5,191, or representative fragments thereof, in a form which can be readily used, analyzed, and interpreted by a skilled artisan.

As used herein, a "representative fragment of the nucleotide sequence depicted in SEQ ID NOS:1-5,191" refers to any portion of the SEQ ID NOS:1-5,191 which is not presently represented within a publicly available database. Preferred representative fragments of the present invention are *Staphylococcus aureus* open reading frames (ORFs), expression modulating fragment (EMFs) and fragments which can be used to diagnose the presence of *Staphylococcus aureus* in sample ("DFs"). A non-limiting identification of preferred representative fragments is provided in Tables 1-3.

As discussed in detail below, the information provided in SEQ ID NOS:1-5,191 and in Tables 1-3 together with routine cloning, synthesis, sequencing and assay methods will enable those skilled in the art to clone and sequence all "representative fragments" of interest, including open reading frames encoding a large variety of *Staphylococcus aureus* proteins.

While the presently disclosed sequences of SEQ ID NOS:1-5,191 are highly accurate, sequencing techniques are not perfect and, in relatively rare instances, further investigation of a fragment or sequence of the invention may reveal a nucleotide sequence error present in a nucleotide sequence disclosed in SEQ ID NOS:1-5,191. However, once the present invention is made available (*i.e.*, once the information in SEQ ID NOS:1-5,191 and Tables 1-3 has been made available), resolving a rare sequencing error in SEQ ID NOS:1-5,191 will be well within the skill of the art. The present disclosure makes available sufficient sequence information to allow any of the described contigs or portions thereof to be obtained readily by straightforward application of routine techniques. Further sequencing of such polynucleotide may proceed in like manner using manual and automated sequencing methods which are employed ubiquitous in the art. Nucleotide sequence editing software is publicly available. For example, Applied Biosystem's (AB) AutoAssembler can be used as an aid during visual inspection of nucleotide sequences. By employing such routine techniques potential errors readily may be identified and the correct sequence then may be ascertained by targeting further sequencing effort, also of a routine nature, to the region containing the potential error.

Even if all of the very rare sequencing errors in SEQ ID NOS:1-5,191 were corrected, the resulting nucleotide sequences would still be at least 95% identical, nearly all would be at least 99% identical, and the great majority would be at least 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-5,191.

As discussed elsewhere hererin, polynucleotides of the present invention readily may be obtained by routine application of well known and standard procedures for cloning and sequencing DNA. Detailed methods for obtaining

libraries and for sequencing are provided below, for instance. A wide variety of *Staphylococcus aureus* strains that can be used to prepare *S. aureus* genomic DNA for cloning and for obtaining polynucleotides of the present invention are available to the public from recognized depository institutions, such as the American Type Culture Collection (ATCC<sup>®</sup>).

5 The nucleotide sequences of the genomes from different strains of *Staphylococcus aureus* differ somewhat. However, the nucleotide sequences of the genomes of all *Staphylococcus aureus* strains will be at least 95% identical, in corresponding part, to the nucleotide sequences provided in SEQ ID NOS:1-5,191. Nearly all will be at least 99% identical and the great majority will be 99.9% identical.

Thus, the present invention further provides nucleotide sequences which are at least 95%, preferably 99% and most preferably 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-5,191, in a form which can be readily  
10 used, analyzed and interpreted by the skilled artisan.

Methods for determining whether a nucleotide sequence is at least 95%, at least 99% or at least 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-5,191 are routine and readily available to the skilled artisan. For example, the well known fasta algorithm described in Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* **65**: 2444 (1988) can be used to generate the percent identity of nucleotide sequences. The BLASTN program also can be used to generate  
15 an identity score of polynucleotides compared to one another.

### COMPUTER RELATED EMBODIMENTS

The nucleotide sequences provided in SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide  
20 sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a polynucleotide sequence of SEQ ID NOS:1-5,191 may be "provided" in a variety of mediums to facilitate use thereof. As used herein, "provided" refers to a manufacture, other than an isolated nucleic acid molecule, which contains a nucleotide sequence of the present invention; i.e., a nucleotide sequence provided in SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical  
25 to a polynucleotide of SEQ ID NOS:1-5,191. Such a manufacture provides a large portion of the *Staphylococcus aureus* genome and parts thereof (e.g., a *Staphylococcus aureus* open reading frame (ORF)) in a form which allows a skilled artisan to examine the manufacture using means not directly applicable to examining the *Staphylococcus aureus* genome or a subset thereof as it exists in nature or in purified form.

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer  
30 readable media. As used herein, "computer readable media" refers to any medium which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories, such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention. Likewise,  
35 it will be clear to those of skill how additional computer readable media that may be developed also can be used to create analogous manufactures having recorded thereon a nucleotide sequence of the present invention.

As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled  
40 artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor  
45 programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data-processor structuring formats (e.g., text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

50 Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. Thus, by providing in computer readable form the nucleotide sequences of SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a sequence of SEQ ID NOS:1-5,191 the present invention enables the skilled artisan routinely to access the provided sequence information for a wide variety of purposes.

55 The examples which follow demonstrate how software which implements the BLAST (Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990)) and BLAZE (Brutlag *et al.*, *Comp. Chem.* 17:203-207 (1993)) search algorithms on a Sybase system was used to identify open reading frames (ORFs) within the *Staphylococcus aureus* genome which contain homology to ORFs or proteins from both *Staphylococcus aureus* and from other organisms. Among the ORFs discussed

herein are protein encoding fragments of the *Staphylococcus aureus* genome useful in producing commercially important proteins, such as enzymes used in fermentation reactions and in the production of commercially useful metabolites.

The present invention further provides systems, particularly computer-based systems, which contain the sequence information described herein. Such systems are designed to identify, among other things, commercially important fragments of the *Staphylococcus aureus* genome.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based system are suitable for use in the present invention.

As stated above, the computer-based systems of the present invention comprise a data storage means having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means.

As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present invention.

As used herein, "search means" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of the present genomic sequences which match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBIA). A skilled artisan can readily recognize that any one of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems.

As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzymic active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

A variety of structural formats for the input and output means can be used to input and output the information in the computer-based systems of the present invention. A preferred format for an output means ranks fragments of the *Staphylococcus aureus* genomic sequences possessing varying degrees of homology to the target sequence or target motif. Such presentation provides a skilled artisan with a ranking of sequences which contain various amounts of the target sequence or target motif and identifies the degree of homology contained in the identified fragment.

A variety of comparing means can be used to compare a target sequence or target motif with the data storage means to identify sequence fragments of the *Staphylococcus aureus* genome. In the present examples, implementing software which implement the BLAST and BLAZE algorithms, described in Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990), was used to identify open reading frames within the *Staphylococcus aureus* genome. A skilled artisan can readily recognize that any one of the publicly available homology search programs can be used as the search means for the computer-based systems of the present invention. Of course, suitable proprietary systems that may be known to those of skill also may be employed in this regard.

Figure 1 provides a block diagram of a computer system illustrative of embodiments of this aspect of present invention. The computer system 102 includes a processor 106 connected to a bus 104. Also connected to the bus 104 are a main memory 108 (preferably implemented as random access memory, RAM) and a variety of secondary storage devices 110, such as a hard drive 112 and a removable medium storage device 114. The removable medium storage device 114 may represent, for example, a floppy disk drive, a CD-ROM drive, a magnetic tape drive, *etc.* A removable storage medium 116 (such as a floppy disk, a compact disk, a magnetic tape, *etc.*) containing control logic and/or data recorded therein may be inserted into the removable medium storage device 114. The computer system 102 includes appropriate software for reading the control logic and/or the data from the removable medium storage device 114, once it is inserted into the removable medium storage device 114.

A nucleotide sequence of the present invention may be stored in a well known manner in the main memory 108, any of the secondary storage devices 110, and/or a removable storage medium 116. During execution, software for

accessing and processing the genomic sequence (such as search tools, comparing tools, etc.) reside in main memory 108, in accordance with the requirements and operating parameters of the operating system, the hardware system and the software program or programs.

## 5 BIOCHEMICAL EMBODIMENTS

Other embodiments of the present invention are directed to fragments of the *Staphylococcus aureus* genome, preferably to isolated fragments. The fragments of the *Staphylococcus aureus* genome of the present invention include, but are not limited to fragments which encode peptides, hereinafter open reading frames (ORFs), fragments which  
10 modulate the expression of an operably linked ORF, hereinafter expression modulating fragments (EMFs) and fragments which can be used to diagnose the presence of *Staphylococcus aureus* in a sample, hereinafter diagnostic fragments (DFs).

As used herein, an "isolated nucleic acid molecule" or an "isolated fragment of the *Staphylococcus aureus* genome" refers to a nucleic acid molecule possessing a specific nucleotide sequence which has been subjected to purification means to reduce, from the composition, the number of compounds which are normally associated with the composition. Particularly, the term refers to the nucleic acid molecules having the sequences set out in SEQ ID NOS:1-5,191, to  
15 representative fragments thereof as described above, to polynucleotides at least 95%, preferably at least 99% and especially preferably at least 99.9% identical in sequence thereto, also as set out above.

A variety of purification means can be used to generate the isolated fragments of the present invention. These  
20 include, but are not limited to methods which separate constituents of a solution based on charge, solubility, or size.

In one embodiment, *Staphylococcus aureus* DNA can be mechanically sheared to produce fragments of 15-20 kb in length. These fragments can then be used to generate an *Staphylococcus aureus* library by inserting them into lambda clones as described in the Examples below. Primers flanking, for example, an ORF, such as those enumerated in Tables 1-3 can then be generated using nucleotide sequence information provided in SEQ ID NOS: 1-5,191. Well  
25 known and routine techniques of PCR cloning then can be used to isolate the ORF from the lambda DNA library of *Staphylococcus aureus* genomic DNA. Thus, given the availability of SEQ ID NOS:1-5,191, the information in Tables 1, 2 and 3, and the information that may be obtained readily by analysis of the sequences of SEQ ID NOS:1-5,191 using methods set out above, those of skill will be enabled by the present disclosure to isolate any ORF-containing or other nucleic acid fragment of the present invention.

The isolated nucleic acid molecules of the present invention include, but are not limited to single stranded and double stranded DNA, and single stranded RNA.

As used herein, an "open reading frame," ORF, means a series of triplets coding for amino acids without any termination codons and is a sequence translatable into protein.

Tables 1, 2 and 3 list ORFs in the *Staphylococcus aureus* genomic contigs of the present invention that were  
35 identified as putative coding regions by the GeneMark software using organism-specific second-order Markov probability transition matrices. It will be appreciated that other criteria can be used, in accordance with well known analytical methods, such as those discussed herein, to generate more inclusive, more restrictive or more selective lists.

Table 1 sets out ORFs in the *Staphylococcus aureus* contigs of the present invention that are at least 80 amino acids long and over a continuous region of at least 50 bases which are 95% or more identical (by BLAST analysis) to an *S. aureus* nucleotide sequence available through Genbank in November 1996.

Table 2 sets out ORFs in the *Staphylococcus aureus* contigs of the present invention that are not in Table 1 and match, with a BLASTP probability score of 0.01 or less, a polypeptide sequence available through Genbank by September 1996.

Table 3 sets out ORFs in the *Staphylococcus aureus* contigs of the present invention that do not match significantly, by BLASTP analysis, a polypeptide sequence available through Genbank by September 1996.

In each table, the first and second columns identify the ORF by, respectively, contig number and ORF number within the contig; the third column indicates the reading frame, taking the first 5' nucleotide of the contig as the start of the +1 frame; the fourth column indicates the first nucleotide of the ORF, counting from the 5' end of the contig strand; and the fifth column indicates the length of each ORF in nucleotides.

In Tables 1 and 2, column six, lists the "Reference" for the closest matching sequence available through Genbank. These reference numbers are the databases entry numbers commonly used by those of skill in the art, who will be familiar with their denominators. Descriptions of the nomenclature are available from the National Center for Biotechnology Information. Column seven in Tables 1 and 2 provides the gene name of the matching sequence; column eight provides the BLAST identity score from the comparison of the ORF and the homologous gene; and column nine  
55 indicates the length in nucleotides of the highest scoring segment pair identified by the BLAST identity analysis.

In Table 3, the last column, column six, indicates the length of each ORF in amino acid residues.

The concepts of percent identity and percent similarity of two polypeptide sequences is well understood in the art. For example, two polypeptides 10 amino acids in length which differ at three amino acid positions (e.g., at positions

1, 3 and 5) are said to have a percent identity of 70%. However, the same two polypeptides would be deemed to have a percent similarity of 80% if, for example at position 5, the amino acids moieties, although not identical, were "similar" (i.e., possessed similar biochemical characteristics). Many programs for analysis of nucleotide or amino acid sequence similarity, such as fasta and BLAST specifically list per cent identity of a matching region as an output parameter. Thus, for instance, Tables 1 and 2 herein enumerate the per cent identity\* of the highest scoring segment pair\* in each ORF and its listed relative. Further details concerning the algorithms and criteria used for homology searches are provided below and are described in the pertinent literature highlighted by the citations provided below.

It will be appreciated that other criteria can be used to generate more inclusive and more exclusive listings of the types set out in the tables. As those of skill will appreciate, narrow and broad searches both are useful. Thus, a skilled artisan can readily identify ORFs in contigs of the *Staphylococcus aureus* genome other than those listed in Tables 1-3, such as ORFs which are overlapping or encoded by the opposite strand of an identified ORF in addition to those ascertainable using the computer-based systems of the present invention.

As used herein, an "expression modulating fragment," EMF, means a series of nucleotide molecules which modulates the expression of an operably linked ORF or EMF.

As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF. EMFs include, but are not limited to, promoters, and promoter modulating sequences (inducible elements). One class of EMFs are fragments which induce the expression or an operably linked ORF in response to a specific regulatory factor or physiological event.

EMF sequences can be identified within the contigs of the *Staphylococcus aureus* genome by their proximity to the ORFs provided in Tables 1-3. An intergenic segment, or a fragment of the intergenic segment, from about 10 to 200 nucleotides in length, taken from any one of the ORFs of Tables 1-3 will modulate the expression of an operably linked ORF in a fashion similar to that found with the naturally linked ORF sequence. As used herein, an "intergenic segment" refers to fragments of the *Staphylococcus aureus* genome which are between two ORF(s) herein described. EMFs also can be identified using known EMFs as a target sequence or target motif in the computer-based systems of the present invention. Further, the two methods can be combined and used together.

The presence and activity of an EMF can be confirmed using an EMF trap vector. An EMF trap vector contains a cloning site linked to a marker sequence. A marker sequence encodes an identifiable phenotype, such as antibiotic resistance or a complementing nutrition auxotrophic factor, which can be identified or assayed when the EMF trap vector is placed within an appropriate host under appropriate conditions. As described above, a EMF will modulate the expression of an operably linked marker sequence. A more detailed discussion of various marker sequences is provided below.

A sequence which is suspected as being an EMF is cloned in all three reading frames in one or more restriction sites upstream from the marker sequence in the EMF trap vector. The vector is then transformed into an appropriate host using known procedures and the phenotype of the transformed host is examined under appropriate conditions. As described above, an EMF will modulate the expression of an operably linked marker sequence.

As used herein, a "diagnostic fragment," DF, means a series of nucleotide molecules which selectively hybridize to *Staphylococcus aureus* sequences. DFs can be readily identified by identifying unique sequences within contigs of the *Staphylococcus aureus* genome, such as by using well-known computer analysis software, and by generating and testing probes or amplification primers consisting of the DF sequence in an appropriate diagnostic format which determines amplification or hybridization selectivity.

The sequences falling within the scope of the present invention are not limited to the specific sequences herein described, but also include allelic and species variations thereof. Allelic and species variations can be routinely determined by comparing the sequences provided in SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably 99% and most preferably 99.9% identical to SEQ ID NOS:1-5,191, with a sequence from another isolate of the same species.

Furthermore, to accommodate codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the nucleic acid sequences mentioned above. In other words, in the coding region of an ORF, substitution of one codon for another which encodes the same amino acid is expressly contemplated.

Any specific sequence disclosed herein can be readily screened for errors by resequencing a particular fragment, such as an ORF, in both directions (i.e., sequence both strands). Alternatively, error screening can be performed by sequencing corresponding polynucleotides of *Staphylococcus aureus* origin isolated by using part or all of the fragments in question as a probe or primer.

Each of the ORFs of the *Staphylococcus aureus* genome disclosed in Tables 1, 2 and 3, and the EMFs found 5' to the ORFs, can be used as polynucleotide reagents in numerous ways. For example, the sequences can be used as diagnostic probes or diagnostic amplification primers to detect the presence of a specific microbe in a sample, particular *Staphylococcus aureus*. Especially preferred in this regard are ORF such as those of Table 3, which do not match previously characterized sequences from other organisms and thus are most likely to be highly selective for *Staphylococcus aureus*. Also particularly preferred are ORFs that can be used to distinguish between strains of *Sta-*

*phylococcus aureus*, particularly those that distinguish medically important strain, such as drug-resistant strains.

In addition, the fragments of the present invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on the binding of a polynucleotide sequence to DNA or RNA. Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Information from the sequences of the present invention can be used to design antisense and triple helix-forming oligonucleotides. Polynucleotides suitable for use in these methods are usually 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription, for triple-helix formation, or to the mRNA itself, for antisense inhibition. Both techniques have been demonstrated to be effective in model systems, and the requisite techniques are well known and involve routine procedures. Triple helix techniques are discussed in, for example, Lee *et al.*, *Nucl. Acids Res.* 6: 3073 (1979); Cooney *et al.*, *Science* 241: 456 (1988); and Dervan *et al.*, *Science* 251: 1360 (1991). Antisense techniques in general are discussed in, for instance, Okano, *J. Neurochem.* 56: 560 (1991) and OLIGODEOXYNUCLEOTIDES AS ANTISENSE INHIBITORS OF GENE EXPRESSION, CRC Press, Boca Raton, FL (1988).

The present invention further provides recombinant constructs comprising one or more fragments of the *Staphylococcus aureus* genomic fragments and contigs of the present invention. Certain preferred recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into which a fragment of the *Staphylococcus aureus* genome has been inserted, in a forward or reverse orientation. In the case of a vector comprising one of the ORFs of the present invention, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the ORF. For vectors comprising the EMFs of the present invention, the vector may further comprise a marker sequence or heterologous ORF operably linked to the EMF.

Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the present invention. The following vectors are provided by way of example. Useful bacterial vectors include phagescript, PsiX174, pBluescript SK and KS (+ and -), pNH8a, pNH16a, pNH18a, pNH46a (available from Stratagene); pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (available from Pharmacia). Useful eukaryotic vectors include pWLneo, pSV2cat, pOG44, pXT1, pSG (available from Stratagene) pSVK3, pBPV, pMSG, pSVL (available from Pharmacia).

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

The present invention further provides host cells containing any one of the isolated fragments of the *Staphylococcus aureus* genomic fragments and contigs of the present invention, wherein the fragment has been introduced into the host cell using known methods. The host cell can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or a procaryotic cell, such as a bacterial cell.

A polynucleotide of the present invention, such as a recombinant construct comprising an ORF of the present invention, may be introduced into the host by a variety of well established techniques that are standard in the art, such as calcium phosphate transfection, DEAE, dextran mediated transfection and electroporation, which are described in, for instance, Davis, L. *et al.*, BASIC METHODS IN MOLECULAR BIOLOGY (1986).

A host cell containing one of the fragments of the *Staphylococcus aureus* genomic fragments and contigs of the present invention, can be used in conventional manners to produce the gene product encoded by the isolated fragment (in the case of an ORF) or can be used to produce a heterologous protein under the control of the EMF.

The present invention further provides isolated polypeptides encoded by the nucleic acid fragments of the present invention or by degenerate variants of the nucleic acid fragments of the present invention. By "degenerate variant" is intended nucleotide fragments which differ from a nucleic acid fragment of the present invention (e.g., an ORF) by nucleotide sequence but, due to the degeneracy of the Genetic Code, encode an identical polypeptide sequence.

Preferred nucleic acid fragments of the present invention are the ORFs depicted in Tables 2 and 3 which encode proteins.

A variety of methodologies known in the art can be utilized to obtain any one of the isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. This is particularly useful in producing small peptides and fragments of larger polypeptides. Such short fragments as may be obtained most readily by synthesis are useful, for example, in generating antibodies against the native polypeptide, as discussed further below.

In an alternative method, the polypeptide or protein is purified from bacterial cells which naturally produce the polypeptide or protein. One skilled in the art can readily employ well-known methods for isolating polypeptides and proteins to isolate and purify polypeptides or proteins of the present invention produced naturally by a bacterial strain, or by other methods. Methods for isolation and purification that can be employed in this regard include, but are not limited to, immunochromatography, HPLC, size-exclusion chromatography, ion-exchange chromatography, and immu-

no-affinity chromatography.

The polypeptides and proteins of the present invention also can be purified from cells which have been altered to express the desired polypeptide or protein. As used herein, a cell is said to be altered to express a desired polypeptide or protein when the cell, through genetic manipulation, is made to produce a polypeptide or protein which it normally does not produce or which the cell normally produces at a lower level. Those skilled in the art can readily adapt procedures for introducing and expressing either recombinant or synthetic sequences into eukaryotic or prokaryotic cells in order to generate a cell which produces one of the polypeptides or proteins of the present invention.

Any host/vector system can be used to express one or more of the ORFs of the present invention. These include, but are not limited to, eukaryotic hosts such as HeLa cells, CV-1 cell, COS cells, and Sf9 cells, as well as prokaryotic host such as *E. coli* and *B. subtilis*. The most preferred cells are those which do not normally express the particular polypeptide or protein or which expresses the polypeptide or protein at low natural level.

"Recombinant," as used herein, means that a polypeptide or protein is derived from recombinant (e.g., microbial or mammalian) expression systems. "Microbial" refers to recombinant polypeptides or proteins made in bacterial or fungal (e.g., yeast) expression systems. As a product, "recombinant microbial" defines a polypeptide or protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Polypeptides or proteins expressed in most bacterial cultures, e.g., *E. coli*, will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern different from that expressed in mammalian cells.

"Nucleotide sequence" refers to a heteropolymer of deoxyribonucleotides. Generally, DNA segments encoding the polypeptides and proteins provided by this invention are assembled from fragments of the *Staphylococcus aureus* genome and short oligonucleotide linkers, or from a series of oligonucleotides, to provide a synthetic gene which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon.

"Recombinant expression vehicle or vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. The expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic regulatory elements necessary for gene expression in the host, including elements required to initiate and maintain transcription at a level sufficient for suitable expression of the desired polypeptide, including, for example, promoters and, where necessary, an enhancers and a polyadenylation signal; (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3) appropriate signals to initiate translation at the beginning of the desired coding region and terminate translation at its end. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an N-terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

"Recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extra chromosomally. The cells can be prokaryotic or eukaryotic. Recombinant expression systems as defined herein will express heterologous polypeptides or proteins upon induction of the regulatory elements linked to the DNA segment or synthetic gene to be expressed.

Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described in Sambrook *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL, 2<sup>nd</sup> Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference in its entirety.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), alpha-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an N-terminal identification peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed recombinant product.

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and, when desirable, provide amplification within the host.

Suitable prokaryotic hosts for transformation include strains of *Staphylococcus aureus*, *E. coli*, *B. subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*. Others

may, also be employed as a matter of choice.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (available from Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (available from Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter, where it is inducible, is derepressed or induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period to provide for expression of the induced gene product. Thereafter cells are typically harvested, generally by centrifugation, disrupted to release expressed protein, generally by physical or chemical means, and the resulting crude extract is retained for further purification.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described in Gluzman, *Cell* **23**: 175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines.

Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

Recombinant polypeptides and proteins produced in bacterial culture is usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

An additional aspect of the invention includes *Staphylococcus aureus* polypeptides which are useful as immunodiagnostic antigens and/or immunoprotective vaccines, collectively "immunologically useful polypeptides". Such immunologically useful polypeptides may be selected from the ORFs disclosed herein based on techniques well known in the art and described elsewhere herein. The inventors have used the following criteria to select several immunologically useful polypeptides:

As is known in the art, an amino terminal type I signal sequence directs a nascent protein across the plasma and outer membranes to the exterior of the bacterial cell. Such outer membrane polypeptides are expected to be immunologically useful. According to Izard, J. W. et al., *Mol. Microbiol.* **13**, 765-773; (1994), polypeptides containing type I signal sequences contain the following physical attributes: The length of the type I signal sequence is approximately 15 to 25 primarily hydrophobic amino acid residues with a net positive charge in the extreme amino terminus; the central region of the signal sequence must adopt an alpha-helical conformation in a hydrophobic environment; and the region surrounding the actual site of cleavage is ideally six residues long, with small side-chain amino acids in the -1 and -3 positions.

Also known in the art is the type IV signal sequence which is an example of the several types of functional signal sequences which exist in addition to the type I signal sequence detailed above. Although functionally related, the type IV signal sequence possesses a unique set of biochemical and physical attributes (Strom, M. S. and Lory, S., *J. Bacteriol.* **174**, 7345-7351; 1992). These are typically six to eight amino acids with a net basic charge followed by an additional sixteen to thirty primarily hydrophobic residues. The cleavage site of a type IV signal sequence is typically after the initial six to eight amino acids at the extreme amino terminus. In addition, all type IV signal sequences contain a phenylalanine residue at the +1 site relative to the cleavage site.

Studies of the cleavage sites of twenty-six bacterial lipoprotein precursors has allowed the definition of a consensus amino acid sequence for lipoprotein cleavage. Nearly three-fourths of the bacterial lipoprotein precursors examined contained the sequence L-(A,S)-(G,A)-C at positions -3 to +1, relative to the point of cleavage (Hayashi, S. and Wu, H. C. *Lipoproteins in bacteria*. *J. Bioenerg. Biomembr.* **22**, 451-471; 1990).

It well known that most anchored proteins found on the surface of gram-positive bacteria possess a highly conserved carboxy terminal sequence. More than fifty such proteins from organisms such as *S. pyogenes*, *S. mutans*, *E. faecalis*, *S. pneumoniae*, and others, have been identified based on their extracellular location and carboxy terminal amino acid sequence (Fischetti, V. A. Gram-positive commensal bacteria deliver antigens to elicit mucosal and systemic immunity. *ASM News* **62**, 405410; 1996). The conserved region is comprised of six charged amino acids at the extreme carboxy terminus coupled to 15-20 hydrophobic amino acids presumed to function as a transmembrane domain. Immediately adjacent to the transmembrane domain is a six amino acid sequence conserved in nearly all proteins ex-

aminated. The amino acid sequence of this region is L-P-X-T-G-X, where X is any amino acid.

Amino acid sequence similarities to proteins of known function by BLAST enables the assignment of putative functions to novel amino acid sequences and allows for the selection of proteins thought to function outside the cell wall. Such proteins are well known in the art and include "lipoprotein", "periplasmic", or "antigen".

An algorithm for selecting antigenic and immunogenic *Staphylococcus aureus* polypeptides including the foregoing criteria was developed by the present inventors. Use of the algorithm by the inventors to select immunologically useful *Staphylococcus aureus* polypeptides resulted in the selection of several ORFs which are predicted to be outer-membrane-associated proteins. These proteins are identified in Table 4, below, and shown in the Sequence Listing as SEQ ID NOS:5,192 to 5,255. Thus the amino acid sequence of each of several antigenic *Staphylococcus aureus* polypeptides listed in Table 4 can be determined, for example, by locating the amino acid sequence of the ORF in the Sequence Listing. Likewise the polynucleotide sequence encoding each ORF can be found by locating the corresponding polynucleotide SEQ ID in Tables 1, 2, or 3, and finding the corresponding nucleotide sequence in the sequence listing.

As will be appreciated by those of ordinary skill in the art, although a polypeptide representing an entire ORF may be the closest approximation to a protein found *in vivo*, it is not always technically practical to express a complete ORF *in vitro*. It may be very challenging to express and purify a highly hydrophobic protein by common laboratory methods. As a result, the immunologically useful polypeptides described herein as SEQ ID NOS:5,192-5,255 may have been modified slightly to simplify the production of recombinant protein, and are the preferred embodiments. In general, nucleotide sequences which encode highly hydrophobic domains, such as those found at the amino terminal signal sequence, are excluded for enhanced *in vitro* expression of the polypeptides. Furthermore, any highly hydrophobic amino acid sequences occurring at the carboxy terminus are also excluded. Such truncated polypeptides include for example the mature forms of the polypeptides expected to exist in nature.

Those of ordinary skill in the art can identify soluble portions the polypeptide identified in Table 4, and in the case of truncated polypeptides sequences shown as SEQ ID NOS:5,192-5,255, may obtain the complete predicted amino acid sequence of each polypeptide by translating the corresponding polynucleotides sequences of the corresponding ORF listed in Tables 1,2 and 3 and found in the sequence listing.

Accordingly, polypeptides comprising the complete amino acid of an immunologically useful polypeptide selected from the group of polypeptides encoded by the ORFs identified in Table 4, or an amino acid sequence at least 95% identical thereto, preferably at least 97% identical thereto, and most preferably at least 99% identical thereto form an embodiment of the invention; in addition polypeptides comprising an amino acid sequence selected from the group of amino acid sequences shown in the sequence listing as SEQ ID NOS:5,191-5,255, or an amino acid sequence at least 95% identical thereto, preferably at least 97% identical thereto and most preferably at least 99% identical thereto, form an embodiment of the invention. Polynucleotides encoding the foregoing polypeptides also form part of the present invention.

In another aspect, the invention provides a peptide or polypeptide comprising an epitope-bearing portion of a polypeptide of the invention, particularly those epitope-bearing portions (antigenic regions) identified in Table 4. The epitope-bearing portion is an immunogenic or antigenic epitope of a polypeptide of the invention. An "immunogenic epitope" is defined as a part of a protein that elicits an antibody response when the whole protein is the immunogen. On the other hand, a region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." The number of immunogenic epitopes of a protein generally is less than the number of antigenic epitopes. See, for instance, Geysen et al., Proc. Natl. Acad. Sci. USA 81:3998- 4002 (1983).

As to the selection of peptides or polypeptides bearing an antigenic epitope (i.e., that contain a region of a protein molecule to which an antibody can bind), it is well known in that art that relatively short synthetic peptides that mimic part of a protein sequence are routinely capable of eliciting an antiserum that reacts with the partially mimicked protein. See, for instance, Sutcliffe, J. G., Shinnick, T. M., Green, N. and Learner, R. A. (1983) "Antibodies that react with predetermined sites on proteins", Science, 219:660-666. Peptides capable of eliciting protein-reactive sera are frequently represented in the primary sequence of a protein, can be characterized by a set of simple chemical rules, and are confined neither to immunodominant regions of intact proteins (i.e., immunogenic epitopes) nor to the amino or carboxyl terminals. Antigenic epitope-bearing peptides and polypeptides of the invention are therefore useful to raise antibodies, including monoclonal antibodies, that bind specifically to a polypeptide of the invention. See, for instance, Wilson et al., Cell 37:767-778 (1984) at 777.

Antigenic epitope-bearing peptides and polypeptides of the invention preferably contain a sequence of at least seven, more preferably at least nine and most preferably between about 15 to about 30 amino acids contained within the amino acid sequence of a polypeptide of the invention. Non-limiting examples of antigenic polypeptides or peptides that can be used to generate *S. aureus* specific antibodies include: a polypeptide comprising peptides shown in Table 4 below. These polypeptide fragments have been determined to bear antigenic epitopes of indicated *S. aureus* proteins by the analysis of the Jameson-Wolf antigenic index, a representative sample of which is shown in Figure 3.

The epitope-bearing peptides and polypeptides of the invention may be produced by any conventional means. See, e.g., Houghten, R. A. (1985) General method for the rapid solid-phase synthesis of large numbers of peptides:

specificity of antigen-antibody interaction at the level of individual amino acids. Proc. Natl. Acad. Sci. USA 82: 5131-5135; this "Simultaneous Multiple Peptide Synthesis (SMPS)" process is further described in U.S. Patent No. 4,631,211 to Houghten et al. (1986). Epitope-bearing peptides and polypeptides of the invention are used to induce antibodies according to methods well known in the art. See, for instance, Sutcliffe et al., supra; Wilson et al., supra; Chow, M. et al., Proc. Natl. Acad. Sci. USA 82:910-914; and Bittle, F. J. et al., J. Gen. Virol. 66:2347-2354 (1985).

Immunogenic epitope-bearing peptides of the invention, i.e., those parts of a protein that elicit an antibody response when the whole protein is the immunogen, are identified according to methods known in the art. See, for instance, Geysen et al., supra. Further still, U.S. Patent No. 5,194,392 to Geysen (1990) describes a general method of detecting or determining the sequence of monomers (amino acids or other compounds) which is a topological equivalent of the epitope (i.e., a "mimotope") which is complementary to a particular paratope (antigen binding site) of an antibody of interest. More generally, U.S. Patent No. 4,433,092 to Geysen (1989) describes a method of detecting or determining a sequence of monomers which is a topographical equivalent of a ligand which is complementary to the ligand binding site of a particular receptor of interest. Similarly, U.S. Patent No. 5,480,971 to Houghten, R. A. et al. (1996) on Per-alkylated Oligopeptide Mixtures discloses linear C1-C7-alkyl peralkylated oligopeptides and sets and libraries of such peptides, as well as methods for using such oligopeptide sets and libraries for determining the sequence of a per-alkylated oligopeptide that preferentially binds to an acceptor molecule of interest. Thus, non-peptide analogs of the epitope-bearing peptides of the invention also can be made routinely by these methods.

Table 4 lists immunologically useful polypeptides identified by an algorithm which locates novel *Staphylococcus aureus* outermembrane proteins, as is described above. Also listed are epitopes or "antigenic regions" of each of the identified polypeptides. The antigenic regions, or epitopes, are delineated by two numbers x-y, where x is the number of the first amino acid in the open reading frame included within the epitope and y is the number of the last amino acid in the open reading frame included within the epitope. For example, the first epitope in ORF 168-6 is comprised of amino acids 36 to 45 of SEQ ID NO:5,192, as is described in Table 4. The inventors have identified several epitopes for each of the antigenic polypeptides identified in Table 4. Accordingly, forming part of the present invention are polypeptides comprising an amino acid sequence of one or more antigenic regions identified in Table 4. The invention further provides polynucleotides encoding such polypeptides.

The present invention further includes isolated polypeptides, proteins and nucleic acid molecules which are substantially equivalent to those herein described. As used herein, substantially equivalent can refer both to nucleic acid and amino acid sequences, for example a mutant sequence, that varies from a reference sequence by one or more substitutions, deletions, or additions, the net effect of which does not result in an adverse functional dissimilarity between reference and subject sequences. For purposes of the present invention, sequences having equivalent biological activity, and equivalent expression characteristics are considered substantially equivalent. For purposes of determining equivalence, truncation of the mature sequence should be disregarded.

The invention further provides methods of obtaining homologs from other strains of *Staphylococcus aureus*, of the fragments of the *Staphylococcus aureus* genome of the present invention and homologs of the proteins encoded by the ORFs of the present invention. As used herein, a sequence or protein of *Staphylococcus aureus* is defined as a homolog of a fragment of the *Staphylococcus aureus* fragments or contigs or a protein encoded by one of the ORFs of the present invention, if it shares significant homology to one of the fragments of the *Staphylococcus aureus* genome of the present invention or a protein encoded by one of the ORFs of the present invention. Specifically, by using the sequence disclosed herein as a probe or as primers, and techniques such as PCR cloning and colony/plaque hybridization, one skilled in the art can obtain homologs.

As used herein, two nucleic acid molecules or proteins are said to "share significant homology" if the two contain regions which possess greater than 85% sequence (amino acid or nucleic acid) homology. Preferred homologs in this regard are those with more than 90% homology. Especially preferred are those with 93% or more homology. Among especially preferred homologs those with 95% or more homology are particularly preferred. Very particularly preferred among these are those with 97% and even more particularly preferred among those are homologs with 99% or more homology. The most preferred homologs among these are those with 99.9% homology or more. It will be understood that, among measures of homology, identity is particularly preferred in this regard.

Region specific primers or probes derived from the nucleotide sequence provided in SEQ ID NOS:1-5,191 or from a nucleotide sequence at least 95%, particularly at least 99%, especially at least 99.5% identical to a sequence of SEQ ID NOS:1-5,191 can be used to prime DNA synthesis and PCR amplification, as well as to identify colonies containing cloned DNA encoding a homolog. Methods suitable to this aspect of the present invention are well known and have been described in great detail in many publications such as, for example, Innis et al., PCR PROTOCOLS, Academic Press, San Diego, CA (1990)).

When using primers derived from SEQ ID NOS:1-5,191 or from a nucleotide sequence having an aforementioned identity to a sequence of SEQ ID NOS:1-5,191, one skilled in the art will recognize that by employing high stringency conditions (e.g., annealing at 50-60°C in 6X SSPE and 50% formamide, and washing at 50-65°C in 0.5X SSPE) only sequences which are greater than 75% homologous to the primer will be amplified. By employing lower stringency

conditions (e.g., hybridizing at 35-37°C in 5X SSPC and 40-45% formamide, and washing at 42°C in 0.5X SSPC), sequences which are greater than 40-50% homologous to the primer will also be amplified.

When using DNA probes derived from SEQ ID NOS:1-5,191, or from a nucleotide sequence having an aforementioned identity to a sequence of SEQ ID NOS:1-5,191, for colony/plaque hybridization, one skilled in the art will recognize that by employing high stringency conditions (e.g., hybridizing at 50-65°C in 5X SSPC and 50% formamide, and washing at 50-65°C in 0.5X SSPC), sequences having regions which are greater than 90% homologous to the probe can be obtained, and that by employing lower stringency conditions (e.g., hybridizing at 35-37°C in 5X SSPC and 40-45% formamide, and washing at 42°C in 0.5X SSPC), sequences having regions which are greater than 35-45% homologous to the probe will be obtained.

Any organism can be used as the source for homologs of the present invention so long as the organism naturally expresses such a protein or contains genes encoding the same. The most preferred organism for isolating homologs are bacteria which are closely related to *Staphylococcus aureus*.

## ILLUSTRATIVE USES OF COMPOSITIONS OF THE INVENTION

Each ORF provided in Tables 1 and 2 is identified with a function by homology to a known gene or polypeptide. As a result, one skilled in the art can use the polypeptides of the present invention for commercial, therapeutic and industrial purposes consistent with the type of putative identification of the polypeptide. Such identifications permit one skilled in the art to use the *Staphylococcus aureus* ORFs in a manner similar to the known type of sequences for which the identification is made; for example, to ferment a particular sugar source or to produce a particular metabolite. A variety of reviews illustrative of this aspect of the invention are available, including the following reviews on the industrial use of enzymes, for example, BIOCHEMICAL ENGINEERING AND BIOTECHNOLOGY HANDBOOK, 2nd Ed., Macmillan Publications, Ltd. NY (1991) and BIOCATALYSTS IN ORGANIC SYNTHESIS, Trumper *et al.*, Eds., Elsevier Science Publishers, Amsterdam, The Netherlands (1985). A variety of exemplary uses that illustrate this and similar aspects of the present invention are discussed below.

### 1. Biosynthetic Enzymes

Open reading frames encoding proteins involved in mediating the catalytic reactions involved in intermediary and macromolecular metabolism, the biosynthesis of small molecules, cellular processes and other functions includes enzymes involved in the degradation of the intermediary products of metabolism, enzymes involved in central intermediary metabolism, enzymes involved in respiration, both aerobic and anaerobic, enzymes involved in fermentation, enzymes involved in ATP proton motor force conversion, enzymes involved in broad regulatory function, enzymes involved in amino acid synthesis, enzymes involved in nucleotide synthesis, enzymes involved in cofactor and vitamin synthesis, can be used for industrial biosynthesis.

The various metabolic pathways present in *Staphylococcus aureus* can be identified based on absolute nutritional requirements as well as by examining the various enzymes identified in Table 1-3 and SEQ ID NOS:1-5,191.

Of particular interest are polypeptides involved in the degradation of intermediary metabolites as well as non-macromolecular metabolism. Such enzymes include amylases, glucose oxidases, and catalase.

Proteolytic enzymes are another class of commercially important enzymes. Proteolytic enzymes find use in a number of industrial processes including the processing of flax and other vegetable fibers, in the extraction, clarification and depectinization of fruit juices, in the extraction of vegetables' oil and in the maceration of fruits and vegetables to give unicellular fruits. A detailed review of the proteolytic enzymes used in the food industry is provided in Rombouts *et al.*, *Symbiosis* 21: 79 (1986) and Voragen *et al.* in BIOCATALYSTS IN AGRICULTURAL BIOTECHNOLOGY, Whitaker *et al.*, Eds., American Chemical Society Symposium Series 389: 93 (1989).

The metabolism of sugars is an important aspect of the primary metabolism of *Staphylococcus aureus*. Enzymes involved in the degradation of sugars, such as, particularly, glucose, galactose, fructose and xylose, can be used in industrial fermentation. Some of the important sugar transforming enzymes, from a commercial viewpoint, include sugar isomerases such as glucose isomerase. Other metabolic enzymes have found commercial use such as glucose oxidases which produces ketogulonic acid (KGA). KGA is an intermediate in the commercial production of ascorbic acid using the Reichstein's procedure, as described in Krueger *et al.*, *Biotechnology* 6(A), Rhine *et al.*, Eds., Verlag Press, Weinheim, Germany (1984).

Glucose oxidase (GOD) is commercially available and has been used in purified form as well as in an immobilized form for the deoxygenation of beer. See, for instance, Hartmeir *et al.*, *Biotechnology Letters* 1: 21 (1979). The most important application of GOD is the industrial scale fermentation of gluconic acid. Market for gluconic acids which are used in the detergent, textile, leather, photographic, pharmaceutical, food, feed and concrete industry, as described, for example, in Bigelis *et al.*, beginning on page 357 in GENE MANIPULATIONS AND FUNGI; Benett *et al.*, Eds., Academic Press, New York (1985). In addition to industrial applications, GOD has found applications in medicine for

quantitative determination of glucose in body fluids recently in biotechnology for analyzing syrups from starch and cellulose hydrosylates. This application is described in Owusu *et al.*, *Biochem. et Biophysica. Acta* 872: 83 (1986), for instance.

The main sweetener used in the world today is sugar which comes from sugar beets and sugar cane. In the field of industrial enzymes, the glucose isomerase process shows the largest expansion in the market today. Initially, soluble enzymes were used and later immobilized enzymes were developed (Krueger *et al.*, *Biotechnology, The Textbook of Industrial Microbiology*, Sinauer Associated Incorporated, Sunderland, Massachusetts (1990)). Today, the use of glucose-produced high fructose syrups is by far the largest industrial business using immobilized enzymes. A review of the industrial use of these enzymes is provided by Jorgensen, *Starch* 40:307 (1988).

Proteinases, such as alkaline serine proteinases, are used as detergent additives and thus represent one of the largest volumes of microbial enzymes used in the industrial sector. Because of their industrial importance, there is a large body of published and unpublished information regarding the use of these enzymes in industrial processes. (See Faultman *et al.*, *Acid Proteinases Structure Function and Biology*, Tang, J., ed., Plenum Press, New York (1977) and Godfrey *et al.*, *Industrial Enzymes*, MacMillan Publishers, Surrey, UK (1983) and Hepner *et al.*, *Report Industrial Enzymes by 1990*, Hel Hepner & Associates, London (1986)).

Another class of commercially usable proteins of the present invention are the microbial lipases, described by, for instance, Macrae *et al.*, *Philosophical Transactions of the Chiral Society of London* 310:227 (1985) and Poserke, *Journal of the American Oil Chemist Society* 61:1758 (1984). A major use of lipases is in the fat and oil industry for the production of neutral glycerides using lipase catalyzed inter-esterification of readily available triglycerides. Application of lipases include the use as a detergent additive to facilitate the removal of fats from fabrics in the course of the washing procedures.

The use of enzymes, and in particular microbial enzymes, as catalyst for key steps in the synthesis of complex organic molecules is gaining popularity at a great rate. One area of great interest is the preparation of chiral intermediates. Preparation of chiral intermediates is of interest to a wide range of synthetic chemists particularly those scientists involved with the preparation of new pharmaceuticals, agrochemicals, fragrances and flavors. (See Davies *et al.*, *Recent Advances in the Generation of Chiral Intermediates Using Enzymes*, CRC Press, Boca Raton, Florida (1990)). The following reactions catalyzed by enzymes are of interest to organic chemists: hydrolysis of carboxylic acid esters, phosphate esters, amides and nitriles, esterification reactions, trans-esterification reactions, synthesis of amides, reduction of alkanones and oxoalkanates, oxidation of alcohols to carbonyl compounds, oxidation of sulfides to sulfoxides, and carbon bond forming reactions such as the aldol reaction.

When considering the use of an enzyme encoded by one of the ORFs of the present invention for biotransformation and organic synthesis it is sometimes necessary to consider the respective advantages and disadvantages of using a microorganism as opposed to an isolated enzyme. Pros and cons of using a whole cell system on the one hand or an isolated partially purified enzyme on the other hand, has been described in detail by Bud *et al.*, *Chemistry in Britain* (1987), p. 127.

Amino transferases, enzymes involved in the biosynthesis and metabolism of amino acids, are useful in the catalytic production of amino acids. The advantages of using microbial based enzyme systems is that the amino transferase enzymes catalyze the stereo-selective synthesis of only L-amino acids and generally possess uniformly high catalytic rates. A description of the use of amino transferases for amino acid production is provided by Roselle-David, *Methods of Enzymology* 136:479 (1987).

Another category of useful proteins encoded by the ORFs of the present invention include enzymes involved in nucleic acid synthesis, repair, and recombination. A variety of commercially important enzymes have previously been isolated from members of *Staphylococcus aureus*. These include Sau3A and Sau96I.

## 2. Generation of Antibodies

As described here, the proteins of the present invention, as well as homologs thereof, can be used in a variety of procedures and methods known in the art which are currently applied to other proteins. The proteins of the present invention can further be used to generate an antibody which selectively binds the protein. Such antibodies can be either monoclonal or polyclonal antibodies, as well fragments of these antibodies, and humanized forms.

The invention further provides antibodies which selectively bind to one of the proteins of the present invention and hybridomas which produce these antibodies. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody.

In general, techniques for preparing polyclonal and monoclonal antibodies as well as hybridomas capable of producing the desired antibody are well known in the art (Campbell, A. M., *MONOCLONAL ANTIBODY TECHNOLOGY: LABORATORY TECHNIQUES IN BIOCHEMISTRY AND MOLECULAR BIOLOGY*, Elsevier Science Publishers, Amsterdam, The Netherlands (1984); St. Groth *et al.*, *J. Immunol. Methods* 35: 1-21 (1980), Kohler and Milstein, *Nature* 256: 495-497 (1975)), the trioma technique, the human B-cell hybridoma technique (Kozbor *et al.*, *Immunology Today*

4: 72 (1983), pgs. 77-96 of Cole *et al.*, in MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc. (1985)).

Any animal (mouse, rabbit, *etc.*) which is known to produce antibodies can be immunized with the pseudogene polypeptide. Methods for immunization are well known in the art. Such methods include subcutaneous or interperitoneal injection of the polypeptide. One skilled in the art will recognize that the amount of the protein encoded by the ORF of the present invention used for immunization will vary based on the animal which is immunized, the antigenicity of the peptide and the site of injection.

The protein which is used as an immunogen may be modified or administered in an adjuvant in order to increase the protein's antigenicity. Methods of increasing the antigenicity of a protein are well known in the art and include, but are not limited to coupling the antigen with a heterologous protein (such as globulin or galactosidase) or through the inclusion of an adjuvant during immunization.

For monoclonal antibodies, spleen cells from the immunized animals are removed, fused with myeloma cells, such as SP2/0-Ag14 myeloma cells, and allowed to become monoclonal antibody producing hybridoma cells.

Any one of a number of methods well known in the art can be used to identify the hybridoma cell which produces an antibody with the desired characteristics. These include screening the hybridomas with an ELISA assay, western blot analysis, or radioimmunoassay (Lutz *et al.*, *Exp. Cell Res.* 175: 109-124 (1988)).

Hybridomas secreting the desired antibodies are cloned and the class and subclass is determined using procedures known in the art (Campbell, A. M., *Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1984)).

Techniques described for the production of single chain antibodies (U. S. Patent 4,946,778) can be adapted to produce single chain antibodies to proteins of the present invention.

For polyclonal antibodies, antibody containing antisera is isolated from the immunized animal and is screened for the presence of antibodies with the desired specificity using one of the above-described procedures.

The present invention further provides the above-described antibodies in detectably labelled form. Antibodies can be detectably labelled through the use of radioisotopes, affinity labels (such as biotin, avidin, *etc.*), enzymatic labels (such as horseradish peroxidase, alkaline phosphatase, *etc.*) fluorescent labels (such as FITC or rhodamine, *etc.*), paramagnetic atoms, *etc.* Procedures for accomplishing such labelling are well-known in the art, for example see Sternberger *et al.*, *J. Histochem. Cytochem.* 18:315 (1970); Bayer, E. A. *et al.*, *Meth. Enzym.* 62:308 (1979); Engval, E. *et al.*, *Immunol.* 109:129 (1972); Goding, J. W. J. *Immunol. Meth.* 13:215 (1976)).

The labeled antibodies of the present invention can be used for *in vitro*, *in vivo*, and in situ assays to identify cells or tissues in which a fragment of the *Staphylococcus aureus* genome is expressed.

The present invention further provides the above-described antibodies immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and sepharose, acrylic resins and such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known in the art (Weir, D. M. *et al.*, "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby, W. D. *et al.*, *Meth. Enzym.* 34 Academic Press, N. Y. (1974)). The immobilized antibodies of the present invention can be used for *in vitro*, *in vivo*, and in situ assays as well as for immunoaffinity purification of the proteins of the present invention.

### 3. Diagnostic Assays and Kits

The present invention further provides methods to identify the expression of one of the ORFs of the present invention, or homolog thereof, in a test sample, using one of the DFs, antigens or antibodies of the present invention.

In detail, such methods comprise incubating a test sample with one or more of the antibodies, or one or more of the DFs, or one or more antigens of the present invention and assaying for binding of the DFs, antigens or antibodies to components within the test sample.

Conditions for incubating a DF, antigen or antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the DF or antibody used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or immunological assay formats can readily be adapted to employ the DFs, antigens or antibodies of the present invention. Examples of such assays can be found in Chard, T., *An Introduction to Radioimmunoassay and Related Techniques*, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G. R. *et al.*, *Techniques in Immunocytochemistry*, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., *Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry*, PCT publication WO95/32291, and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1985), all of which are hereby incorporated herein by reference.

The test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as sputum, blood, serum, plasma, or urine. The test sample used in the above-described method will vary based

on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted in order to obtain a sample which is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the DFs, antigens or antibodies of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound DF, antigen or antibody.

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the antibodies used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound antibody, antigen or DF.

Types of detection reagents include labelled nucleic acid probes, labelled secondary antibodies, or in the alternative, if the primary antibody is labelled, the enzymatic, or antibody binding reagents which are capable of reacting with the labelled antibody. One skilled in the art will readily recognize that the disclosed DFs, antigens and antibodies of the present invention can be readily incorporated into one of the established kit formats which are well known in the art.

#### 4. Screening Assay for Binding Agents

Using the isolated proteins of the present invention, the present invention further provides methods of obtaining and identifying agents which bind to a protein encoded by one of the ORFs of the present invention or to one of the fragments and the *Staphylococcus aureus* fragment and contigs herein described.

In general, such methods comprise steps of:

- (a) contacting an agent with an isolated protein encoded by one of the ORFs of the present invention, or an isolated fragment of the *Staphylococcus aureus* genome; and
- (b) determining whether the agent binds to said protein or said fragment.

The agents screened in the above assay can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. The agents can be selected and screened at random or rationally selected or designed using protein modeling techniques.

For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to the protein encoded by the ORF of the present invention.

Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed" when the agent is chosen based on the configuration of the particular protein. For example, one skilled in the art can readily adapt currently available procedures to generate peptides, pharmaceutical agents and the like capable of binding to a specific peptide sequence in order to generate rationally designed antipeptide peptides, for example see Hurby *et al.*, Application of Synthetic Peptides: Antisense Peptides, "In Synthetic Peptides, A User's Guide, W. H. Freeman, NY (1992), pp. 289-307, and Kaspczak *et al.*, Biochemistry 28:9230-8 (1989), or pharmaceutical agents, or the like.

In addition to the foregoing, one class of agents of the present invention, as broadly described, can be used to control gene expression through binding to one of the ORFs or EMFs of the present invention. As described above, such agents can be randomly screened or rationally designed/selected. Targeting the ORF or EMF allows a skilled artisan to design sequence specific or element specific agents, modulating the expression of either a single ORF or multiple ORFs which rely on the same EMF for expression control.

One class of DNA binding agents are agents which contain base residues which hybridize or form a triple helix by binding to DNA or RNA. Such agents can be based on the classic phosphodiester, ribonucleic acid backbone, or can be a variety of sulfhydryl or polymeric derivatives which have base attachment capacity.

Agents suitable for use in these methods usually contain 20 to 40 bases and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee *et al.*, Nucl. Acids Res. 6:3073 (1979); Cooney *et al.*, Science 241:456 (1988); and Dervan *et al.*, Science 251: 1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated

to be effective in model systems. Information contained in the sequences of the present invention can be used to design antisense and triple helix-forming oligonucleotides, and other DNA binding agents.

## 5. Pharmaceutical Compositions and Vaccines

The present invention further provides pharmaceutical agents which can be used to modulate the growth or pathogenicity of *Staphylococcus aureus*, or another related organism, *in vivo* or *in vitro*. As used herein, a "pharmaceutical agent" is defined as a composition of matter which can be formulated using known techniques to provide a pharmaceutical composition. As used herein, the "pharmaceutical agents of the present invention" refers the pharmaceutical agents which are derived from the proteins encoded by the ORFs of the present invention or are agents which are identified using the herein described assays.

As used herein, a pharmaceutical agent is said to "modulate the growth or pathogenicity of *Staphylococcus aureus* or a related organism, *in vivo* or *in vitro*," when the agent reduces the rate of growth, rate of division, or viability of the organism in question. The pharmaceutical agents of the present invention can modulate the growth or pathogenicity of an organism in many fashions, although an understanding of the underlying mechanism of action is not needed to practice the use of the pharmaceutical agents of the present invention. Some agents will modulate the growth or pathogenicity by binding to an important protein thus blocking the biological activity of the protein, while other agents may bind to a component of the outer surface of the organism blocking attachment or rendering the organism more prone to act the bodies nature immune system. Alternatively, the agent may comprise a protein encoded by one of the ORFs of the present invention and serve as a vaccine. The development and use of vaccines derived from membrane associated polypeptides are well known in the art. The inventors have identified particularly preferred immunogenic *Staphylococcus aureus* polypeptides for use as vaccines. Such immunogenic polypeptides are described above and summarized in Table 4, below.

As used herein, a "related organism" is a broad term which refers to any organism whose growth or pathogenicity can be modulated by one of the pharmaceutical agents of the present invention. In general, such an organism will contain a homolog of the protein which is the target of the pharmaceutical agent or the protein used as a vaccine. As such, related organisms do not need to be bacterial but may be fungal or viral pathogens.

The pharmaceutical agents and compositions of the present invention may be administered in a convenient manner, such as by the oral, topical, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes. The pharmaceutical compositions are administered in an amount which is effective for treating and/or prophylaxis of the specific indication. In general, they are administered in an amount of at least about 1 mg/kg body weight and in most cases they will be administered in an amount not in excess of about 1 g/kg body weight per day. In most cases, the dosage is from about 0.1 mg/kg to about 10 g/kg body weight daily, taking into account the routes of administration, symptoms, etc.

The agents of the present invention can be used in native form or can be modified to form a chemical derivative. As used herein, a molecule is said to be a "chemical derivative" of another molecule when it contains additional chemical moieties not normally a part of the molecule. Such moieties may improve the molecule's solubility, absorption, biological half life, etc. The moieties may alternatively decrease the toxicity of the molecule, eliminate or attenuate any undesirable side effect of the molecule, etc. Moieties capable of mediating such effects are disclosed in, among other sources, REMINGTON'S PHARMACEUTICAL SCIENCES (1980) cited elsewhere herein.

For example, such moieties may change an immunological character of the functional derivative, such as affinity for a given antibody. Such changes in immunomodulation activity are measured by the appropriate assay, such as a competitive type immunoassay. Modifications of such protein properties as redox or thermal stability, biological half-life, hydrophobicity, susceptibility to proteolytic degradation or the tendency to aggregate with carriers or into multimers also may be effected in this way and can be assayed by methods well known to the skilled artisan.

The therapeutic effects of the agents of the present invention may be obtained by providing the agent to a patient by any suitable means (e.g., inhalation, intravenously, intramuscularly, subcutaneously, enterally, or parenterally). It is preferred to administer the agent of the present invention so as to achieve an effective concentration within the blood or tissue in which the growth of the organism is to be controlled. To achieve an effective blood concentration, the preferred method is to administer the agent by injection. The administration may be by continuous infusion, or by single or multiple injections.

In providing a patient with one of the agents of the present invention, the dosage of the administered agent will vary depending upon such factors as the patient's age, weight, height, sex, general medical condition, previous medical history, etc. In general, it is desirable to provide the recipient with a dosage of agent which is in the range of from about 1 pg/kg to 10 mg/kg (body weight of patient), although a lower or higher dosage may be administered. The therapeutically effective dose can be lowered by using combinations of the agents of the present invention or another agent.

As used herein, two or more compounds or agents are said to be administered "in combination" with each other when either (1) the physiological effects of each compound, or (2) the serum concentrations of each compound can

be measured at the same time. The composition of the present invention can be administered concurrently with, prior to, or following the administration of the other agent.

The agents of the present invention are intended to be provided to recipient subjects in an amount sufficient to decrease the rate of growth (as defined above) of the target organism.

The administration of the agent(s) of the invention may be for either a "prophylactic" or "therapeutic" purpose. When provided prophylactically, the agent(s) are provided in advance of any symptoms indicative of the organisms growth. The prophylactic administration of the agent(s) serves to prevent, attenuate, or decrease the rate of onset of any subsequent infection. When provided therapeutically, the agent(s) are provided at (or shortly after) the onset of an indication of infection. The therapeutic administration of the compound(s) serves to attenuate the pathological symptoms of the infection and to increase the rate of recovery.

The agents of the present invention are administered to a subject, such as a mammal, or a patient, in a pharmaceutically acceptable form and in a therapeutically effective concentration. A composition is said to be "pharmacologically acceptable" if its administration can be tolerated by a recipient patient. Such an agent is said to be administered in a "therapeutically effective amount" if the amount administered is physiologically significant. An agent is physiologically significant if its presence results in a detectable change in the physiology of a recipient patient.

The agents of the present invention can be formulated according to known methods to prepare pharmaceutically useful compositions, whereby these materials, or their functional derivatives, are combined in admixture with a pharmaceutically acceptable carrier vehicle. Suitable vehicles and their formulation, inclusive of other human proteins, e. g., human serum albumin, are described, for example, in REMINGTON'S PHARMACEUTICAL SCIENCES, 16<sup>th</sup> Ed., Osol, A., Ed., Mack Publishing, Easton PA (1980). In order to form a pharmaceutically acceptable composition suitable for effective administration, such compositions will contain an effective amount of one or more of the agents of the present invention, together with a suitable amount of carrier vehicle.

Additional pharmaceutical methods may be employed to control the duration of action. Control release preparations may be achieved through the use of polymers to complex or absorb one or more of the agents of the present invention. The controlled delivery may be effectuated by a variety of well known techniques, including formulation with macromolecules such as, for example, polyesters, polyamino acids, polyvinyl, pyrrolidone, ethylenevinylacetate, methylcellulose, carboxymethylcellulose, or protamine, sulfate, adjusting the concentration of the macromolecules and the agent in the formulation, and by appropriate use of methods of incorporation, which can be manipulated to effectuate a desired time course of release. Another possible method to control the duration of action by controlled release preparations is to incorporate agents of the present invention into particles of a polymeric material such as polyesters, polyamino acids, hydrogels, poly(lactic acid) or ethylene vinylacetate copolymers. Alternatively, instead of incorporating these agents into polymeric particles, it is possible to entrap these materials in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization with, for example, hydroxymethylcellulose or gelatine-microcapsules and poly(methylmethacrylate) microcapsules, respectively, or in colloidal drug delivery systems, for example, liposomes, albumin microspheres, microemulsions, nanoparticles, and nanocapsules or in macroemulsions. Such techniques are disclosed in REMINGTON'S PHARMACEUTICAL SCIENCES (1980).

The invention further provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration.

In addition, the agents of the present invention may be employed in conjunction with other therapeutic compounds.

## 6. Shot-Gun Approach to Megabase DNA Sequencing

The present invention further demonstrates that a large sequence can be sequenced using a random shotgun approach. This procedure, described in detail in the examples that follow, has eliminated the up front cost of isolating and ordering overlapping or contiguous subclones prior to the start of the sequencing protocols.

Certain aspects of the present invention are described in greater detail in the examples that follow. The examples are provided by way of illustration. Other aspects and embodiments of the present invention are contemplated by the inventors, as will be clear to those of skill in the art from reading the present disclosure.

ILLUSTRATIVE EXAMPLES

## LIBRARIES AND SEQUENCING

## 5 1. Shotgun Sequencing Probability Analysis

The overall strategy for a shotgun approach to whole genome sequencing follows from the Lander and Waterman (Landerman and Waterman, *Genomics* 2: 231 (1988)) application of the equation for the Poisson distribution. According to this treatment, the probability,  $P_0$ , that any given base in a sequence of size  $L$ , in nucleotides, is not sequenced after a certain amount,  $n$ , in nucleotides, of random sequence has been determined can be calculated by the equation  $P_0 = e^{-m}$ , where  $m$  is  $L/n$ , the fold coverage.\* For instance, for a genome of 2.8 Mb,  $m=1$  when 2.8 Mb of sequence has been randomly generated (1X coverage). At that point,  $P_0 = e^{-1} = 0.37$ . The probability that any given base has not been sequenced is the same as the probability that any region of the whole sequence  $L$  has not been determined and, therefore, is equivalent to the fraction of the whole sequence that has yet to be determined. Thus, at one-fold coverage, approximately 37% of a polynucleotide of size  $L$ , in nucleotides has not been sequenced. When 14 Mb of sequence has been generated, coverage is 5X for a 2.8 Mb and the unsequenced fraction drops to .0067 or 0.67%. 5X coverage of a 2.8 Mb sequence can be attained by sequencing approximately 17,000 random clones from both insert ends with an average sequence read length of 410 bp.

Similarly, the total gap length,  $G$ , is determined by the equation  $G = Le^{-m}$ , and the average gap size,  $g$ , follows the equation,  $g = L/n$ . Thus, 5X coverage leaves about 240 gaps averaging about 82 bp in size in a sequence of a polynucleotide 2.8 Mb long.

The treatment above is essentially that of Lander and Waterman, *Genomics* 2: 231 (1988).

## 2. Random Library Construction

In order to approximate the random model described above during actual sequencing, a nearly ideal library of cloned genomic fragments is required. The following library construction procedure was developed to achieve this end.

*Staphylococcus aureus* DNA was prepared by phenol extraction. A mixture containing 600 ug DNA in 3.3 ml of 300 mM sodium acetate, 10 mM Tris-HCl, 1 mM Na-EDTA, 30% glycerol was sonicated for 1 min. at 0°C in a Branson Model 450 Sonicator at the lowest energy setting using a 3 mm probe. The sonicated DNA was ethanol precipitated and redissolved in 500 ul TE buffer.

To create blunt-ends, a 100 ul aliquot of the resuspended DNA was digested with 5 units of BAL31 nuclease (New England BioLabs) for 10 min at 30°C in 200 ul BAL31 buffer. The digested DNA was phenol-extracted, ethanol-precipitated, redissolved in 100 ul TE buffer, and then size-fractionated by electrophoresis through a 1.0% low melting temperature agarose gel. The section containing DNA fragments 1.6-2.0 kb in size was excised from the gel, and the LGT agarose was melted and the resulting solution was extracted with phenol to separate the agarose from the DNA. DNA was ethanol precipitated and redissolved in 20 ul of TE buffer for ligation to vector.

A two-step ligation procedure was used to produce a plasmid library with 97% inserts, of which >99% were single inserts. The first ligation mixture (50 ul) contained 2 ug of DNA fragments, 2 ug pUC18 DNA (Pharmacia) cut with SmaI and dephosphorylated with bacterial alkaline phosphatase, and 10 units of T4 ligase (GIBCO/BRL) and was incubated at 14°C for 4 hr. The ligation mixture then was phenol extracted and ethanol precipitated, and the precipitated DNA was dissolved in 20 ul TE buffer and electrophoresed on a 1.0% low melting agarose gel. Discrete bands in a ladder were visualized by ethidium bromide-staining and UV illumination and identified by size as insert (i), vector (v), v+i, v+2i, v+3i, etc. The portion of the gel containing v+i DNA was excised and the v+i DNA was recovered and resuspended into 20 ul TE. The v+i DNA then was blunt-ended by T4 polymerase treatment for 5 min. at 37° C in a reaction mixture (50 ul) containing the v+i linears, 500 uM each of the 4 dNTPs, and 9 units of T4 polymerase (New England BioLabs), under recommended buffer conditions. After phenol extraction and ethanol precipitation the repaired v+i linears were dissolved in 20 ul TE. The final ligation to produce circles was carried out in a 50 ul reaction containing 5 ul of v+i linears and 5 units of T4 ligase at 14°C overnight. After 10 min. at 70°C the following day, the reaction mixture was stored at -20°C.

This two-stage procedure resulted in a molecularly random collection of single-insert plasmid recombinants with minimal contamination from double-insert chimeras (<1%) or free vector (<3%).

Since deviation from randomness can arise from propagation the DNA in the host, *E.coli* host cells deficient in all recombination and restriction functions (A. Greener, *Strategies* 3 (1):5 (1990)) were used to prevent rearrangements, deletions, and loss of clones by restriction. Furthermore, transformed cells were plated directly on antibiotic diffusion plates to avoid the usual broth recovery phase which allows multiplication and selection of the most rapidly growing cells.

Plating was carried out as follows. A 100 ul aliquot of Epicurian Coli SURE II Supercompetent Cells (Stratagene 200152) was thawed on ice and transferred to a chilled Falcon 2059 tube on ice. A 1.7 ul aliquot of 1.42 M beta-

mercaptoethanol was added to the aliquot of cells to a final concentration of 25 mM. Cells were incubated on ice for 10 min. A 1  $\mu$ l aliquot of the final ligation was added to the cells and incubated on ice for 30 min. The cells were heat pulsed for 30 sec. at 42° C and placed back on ice for 2 min. The outgrowth period in liquid culture was eliminated from this protocol in order to minimize the preferential growth of any given transformed cell. Instead the transformation mixture was plated directly on a nutrient rich SOB plate containing a 5 ml bottom layer of SOB agar (5% SOB agar: 20 g tryptone, 5 g yeast extract, 0.5 g NaCl, 1.5% Difco Agar per liter of media). The 5 ml bottom layer is supplemented with 0.4 ml of 50 mg/ml ampicillin per 100 ml SOB agar. The 15 ml top layer of SOB agar is supplemented with 1 ml X-Gal (2%), 1 ml  $MgCl_2$  (1 M), and 1 ml  $MgSO_4$ /100 ml SOB agar. The 15 ml top layer was poured just prior to plating. Our titer was approximately 100 colonies/10  $\mu$ l aliquot of transformation.

All colonies were picked for template preparation regardless of size. Thus, only clones lost due to "poison" DNA or deleterious gene products would be deleted from the library, resulting in a slight increase in gap number over that expected.

### 3. Random DNA Sequencing

High quality double stranded DNA plasmid templates were prepared using an alkaline lysis method developed in collaboration with 5Prime  $\rightarrow$  3Prime Inc. (Boulder, CO). Plasmid preparation was performed in a 96-well format for all stages of DNA preparation from bacterial growth through final DNA purification. Average template concentration was determined by running 25% of the samples on an agarose gel. DNA concentrations were not adjusted.

Templates were also prepared from a *Staphylococcus aureus* lambda genomic library. An unamplified library was constructed in Lambda DASH II vector (Stratagene). *Staphylococcus aureus* DNA (> 100 kb) was partially digested in a reaction mixture (200  $\mu$ l) containing 50  $\mu$ g DNA, 1X Sau3AI buffer, 20 units Sau3AI for 6 min. at 23 C. The digested DNA was phenol-extracted and centrifuges over a 10- 40% sucrose gradient. Fractions containing genomic DNA of 15-25 kb were recovered by precipitation. One  $\mu$ l of fragments was used with 1  $\mu$ l of DASHII vector (Stratagene) in the recommended ligation reaction. One  $\mu$ l of the ligation mixture was used per packaging reaction following the recommended protocol with the Gigapack II XL Packaging Extract Phage were plated directly without amplification from the packaging mixture (after dilution with 500  $\mu$ l of recommended SM buffer and chloroform treatment). Yield was about  $2.5 \times 10^9$  pfu/ $\mu$ l.

An amplified library was prepared from the primary packaging mixture according to the manufacturer's protocol. The amplified library is stored frozen in 7% dimethylsulfoxide. The phage titer is approximately  $1 \times 10^9$  pfu/ml.

Mini-liquid lysates (0.1  $\mu$ l) are prepared from randomly selected plaques and template is prepared by long range PCR. Samples are PCR amplified using modified T3 and T7 primers, and Elongase Supermix (LTI).

Sequencing reactions are carried out on plasmid templates using a combination of two workstations (BIOMEK 1000 and Hamilton Microlab 2200) and the Perkin-Elmer 9600 thermocycler with Applied Biosystems PRISM Ready Reaction Dye Primer Cycle Sequencing Kits for the M13 forward (M13-21) and the M13 reverse (M13RP1) primers. Dye terminator sequencing reactions are carried out on the lambda templates on a Perkin-Elmer 9600 Thermocycler using the Applied Biosystems Ready Reaction Dye Terminator Cycle Sequencing kits. Modified T7 and T3 primers are used to sequence the ends of the inserts from the Lambda DASH II library. Sequencing reactions are on a combination of AB 373 DNA Sequencers and ABI 377 DNA sequencers. All of the dye terminator sequencing reactions are analyzed using the 2X 9 hour module on the AB 377. Dye primer reactions are analyzed on a combination of ABI 373 and ABI 377 DNA sequencers. The overall sequencing success rate very approximately is about 85% for M13-21 and M13RP1 sequences and 65% for dye-terminator reactions. The average usable read length is 485 bp for M13-21 sequences, 445bp for M13RP1 sequences, and 375 bp for dye-terminator reactions.

### 4. Protocol for Automated Cycle Sequencing

The sequencing was carried out using Hamilton Microstation 2200, Perkin Elmer 9600 thermocyclers, ABI 373 and ABI 377 Automated DNA Sequencers. The Hamilton combines pre-aliquoted templates and reaction mixes consisting of deoxy- and dideoxynucleotides, the thermostable Taq DNA polymerase, fluorescently-labelled sequencing primers, and reaction buffer. Reaction mixes and templates were combined in the wells of a 96-well thermocycling plate and transferred to the Perkin Elmer 9600 thermocycler. Thirty consecutive cycles of linear amplification (i.e., one primer synthesis) steps were performed including denaturation, annealing of primer and template, and extension; i.e., DNA synthesis. A heated lid with rubber gaskets on the thermocycling plate prevents evaporation without the need for an oil overlay.

Two sequencing protocols were used: one for dye-labelled primers and a second for dye-labelled dideoxy chain terminators. The shotgun sequencing involves use of four dye-labelled sequencing primers, one for each of the four terminator nucleotide. Each dye-primer was labelled with a different fluorescent dye, permitting the four individual reactions to be combined into one lane of the 373 or 377 DNA Sequencer for electrophoresis, detection, and base-

calling. ABI currently supplies premixed reaction mixes in bulk packages containing all the necessary non-template reagents for sequencing. Sequencing can be done with both plasmid and PCR-generated templates with both dye-primers and dye-terminators with approximately equal fidelity, although plasmid templates generally give longer usable sequences.

Thirty-two reactions were loaded per ABI 373 Sequencer each day and 96 samples can be loaded on an ABI 377 per day. Electrophoresis was run overnight (ABI 373) or for 2 1/2 hours (ABI 377) following the manufacturer's protocols. Following electrophoresis and fluorescence detection, the ABI 373 or ABI 377 performs automatic lane tracking and base-calling. The lane-tracking was confirmed visually. Each sequence electropherogram (or fluorescence lane trace) was inspected visually and assessed for quality. Trailing sequences of low quality were removed and the sequence itself was loaded via software to a Sybase database (archived daily to 8mm tape). Leading vector polylinker sequence was removed automatically by a software program. Average edited lengths of sequences from the standard ABI 373 or ABI 377 were around 400 bp and depend mostly on the quality of the template used for the sequencing reaction.

## INFORMATICS

### 1. Data Management

A number of information management systems for a large-scale sequencing lab have been developed. (For review see, for instance, Kerlavage *et al.*, *Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System Sciences*, IEEE Computer Society Press, Washington D. C., 585 (1993)) The system used to collect and assemble the sequence data was developed using the Sybase relational database management system and was designed to automate data flow wherever possible and to reduce user error. The database stores and correlates all information collected during the entire operation from template preparation to final analysis of the genome. Because the raw output of the ABI 373 Sequencers was based on a Macintosh platform and the data management system chosen was based on a Unix platform, it was necessary to design and implement a variety of multi-user, client-server applications which allow the raw data as well as analysis results to flow seamlessly into the database with a minimum of user effort.

### 2. Assembly

An assembly engine (TIGR Assembler) developed for the rapid and accurate assembly of thousands of sequence fragments was employed to generate contigs. The TIGR assembler simultaneously clusters and assembles fragments of the genome. In order to obtain the speed necessary to assemble more than  $10^4$  fragments, the algorithm builds a hash table of 12 bp oligonucleotide subsequences to generate a list of potential sequence fragment overlaps. The number of potential overlaps for each fragment determines which fragments are likely to fall into repetitive elements. Beginning with a single seed sequence fragment, TIGR Assembler extends the current contig by attempting to add the best matching fragment based on oligonucleotide content. The contig and candidate fragment are aligned using a modified version of the Smith-Waterman algorithm which provides for optimal gapped alignments (Waterman, M. S., *Methods in Enzymology* 164: 765 (1988)). The contig is extended by the fragment only if strict criteria for the quality of the match are met. The match criteria include the minimum length of overlap, the maximum length of an unmatched end, and the minimum percentage match. These criteria are automatically lowered by the algorithm in regions of minimal coverage and raised in regions with a possible repetitive element. The number of potential overlaps for each fragment determines which fragments are likely to fall into repetitive elements. Fragments representing the boundaries of repetitive elements and potentially chimeric fragments are often rejected based on partial mismatches at the ends of alignments and excluded from the current contig. TIGR Assembler is designed to take advantage of clone size information coupled with sequencing from both ends of each template. It enforces the constraint that sequence fragments from two ends of the same template point toward one another in the contig and are located within a certain ranged of base pairs (definable for each clone based on the known clone size range for a given library).

### 3. Identifying Genes

The predicted coding regions of the *Staphylococcus aureus* genome were initially defined with the program zorf, which finds ORFs of a minimum length. The predicted coding region sequences were used in searches against a database of all *Staphylococcus aureus* nucleotide sequences from GenBank (release 92.0), using the BLASTN search method to identify overlaps of 50 or more nucleotides with at least a 95% identity. Those ORFs with nucleotide sequence matches are shown in Table 1. The ORFs without such matches were translated to protein sequences and compared to a non-redundant database of known proteins generated by combining the Swiss-prot, PIR and GenPept databases. ORFs of at least 80 amino acids that matched a database protein with BLASTP probability less than or equal to 0.01 are shown in Table 2. The table also lists assigned functions based on the closest match in the databases.

ORFs of at least 120 amino acids that did not match protein or nucleotide sequences in the databases at these levels are shown in Table 3.

## ILLUSTRATIVE APPLICATIONS

### 1. Production of an Antibody to a *Staphylococcus aureus* Protein

Substantially pure protein or polypeptide is isolated from the transfected or transformed cells using any one of the methods known in the art. The protein can also be produced in a recombinant prokaryotic expression system, such as *E. coli*, or can be chemically synthesized. Concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms/ml. Monoclonal or polyclonal antibody to the protein can then be prepared as follows.

### 2. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes of any of the peptides identified and isolated as described can be prepared from murine hybridomas according to the classical method of Kohler, G. and Milstein, C., *Nature* 256:495 (1975) or modifications of the methods thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the selected protein over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, E., *Meth. Enzymol.* 70:419 (1980), and modified methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. *et al.* *Basic Methods in Molecular Biology* Elsevier, New York. Section 21-2 (1989).

### 3. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogenous epitopes of a single protein can be prepared by immunizing suitable animals with the expressed protein described above, which can be unmodified or modified to enhance immunogenicity. Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. For example, small molecules tend to be less immunogenic than other and may require the use of carriers and adjuvant. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. An effective immunization protocol for rabbits can be found in Vaitukaitis, J. *et al.*, *J. Clin. Endocrinol. Metab.* 33:988-991 (1971).

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, O. *et al.*, Chap. 19 in: *Handbook of Experimental Immunology*, Wier, D., ed, Blackwell (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12M). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher, D., Chap. 42 in: *Manual of Clinical Immunology*, second edition, Rose and Friedman, eds., Amer. Soc. For Microbiology, Washington, D. C. (1980).

Antibody preparations prepared according to either protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample. In addition, they are useful in various animal models of *Staphylococcal* disease known to those of skill in the art as a means of evaluating the protein used to make the antibody as a potential vaccine target or as a means of evaluating the antibody as a potential immunotherapeutic reagent.

### 3. Preparation of PCR Primers and Amplification of DNA

Various fragments of the *Staphylococcus aureus* genome, such as those of Tables 1-3 and SEQ ID NOS:1-5,191 can be used, in accordance with the present invention, to prepare PCR primers for a variety of uses. The PCR primers are preferably at least 15 bases, and more preferably at least 18 bases in length. When selecting a primer sequence, it is preferred that the primer pairs have approximately the same G/C ratio, so that melting temperatures are approxi-

mately the same. The PCR primers and amplified DNA of this Example find use in the Examples that follow.

#### 4. Gene expression from DNA Sequences Corresponding to ORFs

A fragment of the *Staphylococcus aureus* genome provided in Tables 1-3 is introduced into an expression vector using conventional technology. Techniques to transfer cloned sequences into expression vectors that direct protein translation in mammalian, yeast, insect or bacterial expression systems are well known in the art. Commercially available vectors and expression systems are available from a variety of suppliers including Stratagene (La Jolla, California), Promega (Madison, Wisconsin), and Invitrogen (San Diego, California). If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence may be optimized for the particular expression organism, as explained by Hatfield *et al.*, U. S. Patent No. 5,082,767, incorporated herein by this reference.

The following is provided as one exemplary method to generate polypeptide(s) from cloned ORFs of the *Staphylococcus aureus* genome fragment. Bacterial ORFs generally lack a poly A addition signal. The addition signal sequence can be added to the construct by, for example, splicing out the poly A addition sequence from pSG5 (Stratagene) using BglI and SalI restriction endonuclease enzymes and incorporating it into the mammalian expression vector pXT1 (Stratagene) for use in eukaryotic expression systems. pXT1 contains the LTRs and a portion of the gag gene of Moloney Murine Leukemia Virus. The positions of the LTRs in the construct allow efficient stable transfection. The vector includes the Herpes Simplex thymidine kinase promoter and the selectable neomycin gene. The *Staphylococcus aureus* DNA is obtained by PCR from the bacterial vector using oligonucleotide primers complementary to the *Staphylococcus aureus* DNA and containing restriction endonuclease sequences for PstI incorporated into the 5' primer and BglII at the 3' end of the corresponding *Staphylococcus aureus* DNA 3' primer, taking care to ensure that the *Staphylococcus aureus* DNA is positioned such that it's followed with the poly A addition sequence. The purified fragment obtained from the resulting PCR reaction is digested with PstI, blunt ended with an exonuclease, digested with BglII, purified and ligated to pXT1, now containing a poly A addition sequence and digested BglII.

The ligated product is transfected into mouse NIH 3T3 cells using Lipofectin (Life Technologies, Inc., Grand Island, New York) under conditions outlined in the product specification. Positive transfectants are selected after growing the transfected cells in 600 ug/ml G418 (Sigma, St. Louis, Missouri). The protein is preferably released into the supernatant. However if the protein has membrane binding domains, the protein may additionally be retained within the cell or expression may be restricted to the cell surface. Since it may be necessary to purify and locate the transfected product, synthetic 15-mer peptides synthesized from the predicted *Staphylococcus aureus* DNA sequence are injected into mice to generate antibody to the polypeptide encoded by the *Staphylococcus aureus* DNA.

Alternatively and if antibody production is not possible, the *Staphylococcus aureus* DNA sequence is additionally incorporated into eukaryotic expression vectors and expressed as, for example, a globin fusion. Antibody to the globin moiety then is used to purify the chimeric protein. Corresponding protease cleavage sites are engineered between the globin moiety and the polypeptide encoded by the *Staphylococcus aureus* DNA so that the latter may be freed from the formed by simple protease digestion. One useful expression vector for generating globin chimerics is pSG5 (Stratagene). This vector encodes a rabbit globin. Intron II of the rabbit globin gene facilitates splicing of the expressed transcript, and the polyadenylation signal incorporated into the construct increases the level of expression. These techniques are well known to those skilled in the art of molecular biology. Standard methods are published in methods texts such as Davis *et al.*, cited elsewhere herein, and many of the methods are available from the technical assistance representatives from Stratagene, Life Technologies, Inc., or Promega. Polypeptides of the invention also may be produced using *in vitro* translation systems such as *in vitro* Express<sup>TM</sup> Translation Kit (Stratagene).

While the present invention has been described in some detail for purposes of clarity and understanding, one skilled in the art will appreciate that various changes in form and detail can be made without departing from the true scope of the invention.

All patents, patent applications and publications referred to above are hereby incorporated by reference.

S. aureus - Coding regions containing known sequences

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | Percent Ident | HSP nt length | ORF nt length |
|-----------|--------|------------|-----------|-----------------|---|---------------|---------------|---------------|
| 1         | 1      | 1419       | 757       | emb X17301 SAMD | S. aureus DNA for hid gene and for part of agr gene   | 100           | 663           | 663           |
| 1         | 2      | 3273       | 2453      | emb X52543 SAG  | S. aureus agrA, agrB and hid genes  | 99            | 809           | 822           |
| 1         | 5      | 6418       | 5651      | dbj D14711 STAM | Staphylococcus aureus HSP10 and HSP60 genes   | 98            | 223           | 768           |
| 5         | 1      | 807        | 439       | emb X72700 SAPV | S. aureus genes for S and F components of Pantone-Valentine leucocidins   | 81            | 216           | 369           |
| 5         | 4      | 5031       | 3571      | emb X72700 SAPV | S. aureus genes for S and F components of Pantone-Valentine leucocidins   | 95            | 424           | 1461          |
| 10        | 1      | 86         | 904       | gb U25288       | Staphylococcus aureus gyrase-like protein alpha and beta subunit (gria and griB) genes, complete cds              | 98            | 715           | 819           |
| 16        | 5      | 5302       | 6246      | gb U35773       | Staphylococcus aureus prolipoprotein diacylglycerol transferase (lgt) gene, complete cds                          | 94            | 251           | 945           |
| 16        | 6      | 6249       | 7091      | gb U35773       | Staphylococcus aureus prolipoprotein diacylglycerol transferase (lgt) gene, complete cds                          | 99            | 843           | 843           |
| 16        | 7      | 7084       | 7584      | gb U35773       | Staphylococcus aureus prolipoprotein diacylglycerol transferase (lgt) gene, complete cds                          | 99            | 342           | 501           |
| 20        | 1      | 995        | 549       | gb U19300       | Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank | 100           | 443           | 447           |
| 20        | 2      | 1011       | 841       | gb U19300       | Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank | 91            | 137           | 171           |
| 20        | 3      | 2010       | 1798      | gb U19300       | Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank | 100           | 110           | 213           |
| 20        | 4      | 5300       | 3825      | gb H16714       | Staphylococcus aureus peptidoglycan hydrolase gene, complete cds  | 100           | 948           | 1476          |
| 20        | 5      | 4788       | 4282      | gb H16714       | Staphylococcus aureus peptidoglycan hydrolase gene, complete cds  | 100           | 309           | 507           |
| 26        | 1      | 2          | 145       | gb U41072       | Staphylococcus aureus isoleucyl-tRNA synthetase (lles) gene, partial cds  | 100           | 126           | 144           |
| 26        | 2      | 84         | 557       | gb U41072       | Staphylococcus aureus isoleucyl-tRNA synthetase (lles) gene, partial cds  | 99            | 430           | 474           |
| 26        | 3      | 763        | 3531      | emb X74219 SAIL | S. aureus gene for isoleucyl-tRNA synthetase  | 99            | 2769          | 2769          |
| 29        | 3      | 1261       | 4392      | gb U66665       | Staphylococcus aureus DNA fragment with class II promoter activity  | 100           | 117           | 3132          |
| 31        | 14     | 14977      | 13463     | emb X73889 SAP1 | S. aureus genes P1 and P2   | 99            | 1351          | 1515          |
| 31        | 15     | 14241      | 13655     | emb X73889 SAP1 | S. aureus genes P1 and P2   | 98            | 258           | 387           |
| 38        | 17     | 14284      | 13112     | gb H12715       | S. aureus geh gene encoding lipase (glycerol ester hydrolase)   | 100           | 372           | 1173          |
| 38        | 19     | 13434      | 15518     | gb P12715       | S. aureus geh gene encoding lipase (glycerol ester hydrolase)   | 100           | 2085          | 2085          |

TABLE I

S. aureus - Coding regions containing known sequences

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name  | Percent ident | HSP nt length | ORF nt length |
|-----------|--------|------------|-----------|-----------------|--|---------------|---------------|---------------|
| 46        | 3      | 519        | 1727      | gb U73374       | Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds | 98            | 1209          | 1209          |
| 46        | 3      | 1720       | 2295      | gb U73374       | Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds | 98            | 576           | 576           |
| 46        | 4      | 2259       | 3182      | gb U73374       | Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds | 97            | 924           | 924           |
| 46        | 5      | 3173       | 4498      | gb U73374       | Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds | 98            | 1283          | 1326          |
| 46        | 6      | 4536       | 5720      | gb U73374       | Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds | 98            | 1185          | 1185          |
| 46        | 7      | 6455       | 6120      | gb U73374       | Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds | 99            | 278           | 336           |
| 48        | 1      | 2          | 955       | gb L25893       | Staphylococcus aureus recA gene, complete cds  | 99            | 954           | 954           |
| 50        | 3      | 4465       | 2924      | emb X85029 SAAH | S. aureus AhpC gene  | 100           | 88            | 1542          |
| 50        | 4      | 4108       | 3515      | emb X85029 SAAH | S. aureus AhpC gene  | 98            | 540           | 594           |
| 54        | 3      | 5074       | 3392      | emb X62992 SAFN | S. aureus fnbB gene for fibronectin binding protein B  | 100           | 1668          | 1683          |
| 54        | 4      | 4865       | 4122      | emb X62992 SAFN | S. aureus fnbB gene for fibronectin binding protein B  | 99            | 720           | 744           |
| 54        | 5      | 5056       | 4562      | emb X62992 SAFN | S. aureus fnbB gene for fibronectin binding protein B  | 100           | 463           | 495           |
| 54        | 6      | 11386      | 8300      | gb J04151       | S. aureus fibronectin-binding protein (fnbA) mRNA, complete cds  | 100           | 3087          | 3087          |
| 58        | 3      | 1743       | 2819      | emb X87104 SADK | S. aureus mdr, pbp4 and tagD genes (SC511-55 isolate)  | 89            | 68            | 1077          |
| 58        | 4      | 2858       | 3280      | emb X91786 SAPB | S. aureus abcA, pbp4, and tagD genes   | 99            | 423           | 423           |
| 58        | 5      | 6005       | 4701      | emb X91786 SAPB | S. aureus abcA, pbp4, and tagD genes   | 99            | 1305          | 1305          |
| 58        | 6      | 5677       | 5378      | gb U29478       | Staphylococcus aureus ABC transporter-like protein AbcA (abcA) gene, partial cds   | 100           | 300           | 300           |
| 58        | 7      | 5086       | 6840      | emb X91786 SAPB | S. aureus abcA, pbp4, and tagD genes   | 99            | 1755          | 1755          |
| 72        | 1      | 888        | 445       | gb U1854        | S. aureus agr gene encoding an accessory gene regulator protein, complete cds  | 100           | 444           | 444           |
| 72        | 2      | 2457       | 1453      | emb X52543 SAGC | S. aureus agrA, agrB and hid genes   | 99            | 673           | 1005          |

TABLE 1

S. aureus - Coding regions containing known sequences

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name  | Percent ident | MSP nt length | ORF nt length |
|-----------|--------|------------|-----------|-----------------|--|---------------|---------------|---------------|
| 82        | 1      | 357        | 3917      | emb X64172 SARP | S. aureus rplL, orf202, rps31(rif) and rpsC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains             | 99            | 2396          | 3561          |
| 82        | 2      | 4027       | 7677      | emb X69233 SARP | S. aureus DNA for rpsC gene  | 99            | 3171          | 3651          |
| 82        | 3      | 7745       | 8068      | gb U20869       | Staphylococcus aureus ribosomal protein S12 (rpsL) gene, complete cds, ribosomal protein S7 (rpsG) and ORF 1 genes, partial cds  | 100           | 320           | 324           |
| 82        | 4      | 8103       | 8579      | gb U20869       | Staphylococcus aureus ribosomal protein S12 (rpsL) gene, complete cds, ribosomal protein S7 (rpsG) and ORF 1 genes, partial cds  | 100           | 477           | 477           |
| 82        | 5      | 8618       | 8821      | gb U20869       | Staphylococcus aureus ribosomal protein S12 (rpsL) gene, complete cds, ribosomal protein S7 (rpsG) and ORF 1 genes, partial cds  | 100           | 154           | 204           |
| 84        | 1      | 18         | 191       | gb U73374       | Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds | 98            | 164           | 174           |
| 84        | 2      | 189        | 893       | gb U73374       | Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds | 94            | 705           | 705           |
| 84        | 3      | 887        | 1660      | gb U73374       | Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds | 99            | 774           | 774           |
| 84        | 4      | 1584       | 3503      | gb U73374       | Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds | 98            | 1920          | 1920          |
| 84        | 5      | 3394       | 4521      | gb U73374       | Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds | 97            | 1128          | 1128          |
| 84        | 6      | 4519       | 5643      | gb U73374       | Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds | 97            | 1125          | 1125          |
| 96        | 2      | 1245       | 3896      | emb Z18852 SACF | S. aureus gene for clumping factor   | 83            | 660           | 2652          |
| 97        | 2      | 625        | 882       | gb U41072       | Staphylococcus aureus isoleucyl-tRNA synthetase (ileS) gene, partial cds   | 97            | 68            | 258           |
| 111       | 1      | 3          | 452       | gb L41499       | Staphylococcus aureus ORF1, partial cds, ORF2, ORF3, autolysin (atl) genes, complete cds   | 100           | 450           | 450           |
| 111       | 2      | 526        | 1041      | gb L41499       | Staphylococcus aureus ORF1, partial cds, ORF2, ORF3, autolysin (atl) genes, complete cds   | 99            | 516           | 516           |
| 117       | 2      | 1278       | 1958      | gb M83994       | Staphylococcus aureus prolipoprotein signal peptidase (lsp) gene, complete cds   | 100           | 61            | 681           |

TABLE 1

S. aureus - Coding regions containing known sequences

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | percent ident | HSP nt length | ORF nt length |
|-----------|--------|------------|-----------|-----------------|---|---------------|---------------|---------------|
| 118       | 4      | 3787       | 4254      | dbj U06990 STAN | Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP10; ORF15, complete cds   | 99            | 467           | 468           |
| 130       | 4      | 2597       | 3640      | emb X13290 SATN | Staphylococcus aureus multi-resistance plasmid pSK1 DNA containing transposon Tn4003  | 78            | 956           | 1044          |
| 130       | 5      | 3813       | 4265      | emb Z16422 SAOI | S. aureus dfrB gene for dihydrofolate reductase   | 98            | 416           | 453           |
| 130       | 6      | 4309       | 5172      | emb Z16422 SAOI | S. aureus dfrB gene for dihydrofolate reductase   | 98            | 607           | 864           |
| 136       | 4      | 5296       | 6207      | emb X71437 SAOY | S. aureus genes gyrB, gyrA and recF (partial)   | 97            | 838           | 912           |
| 136       | 5      | 11680      | 8987      | dbj D10489 STAG | Staphylococcus aureus genes for DNA gyrase A and B, complete cds  | 100           | 2694          | 2694          |
| 136       | 6      | 12886      | 10940     | dbj D10489 STAG | Staphylococcus aureus genes for DNA gyrase A and B, complete cds  | 99            | 1947          | 1947          |
| 136       | 7      | 12592      | 11765     | gb S77055       | recF cluster: dnaA-replisome assembly protein...gyrB-DNA gyrase beta subunit [Staphylococcus aureus, Y8886, Genomic, 3573 nt]           | 99            | 822           | 828           |
| 143       | 3      | 4171       | 2867      | gb U36379       | Staphylococcus aureus S-adenosylmethionine synthetase gene, complete cds  | 99            | 1305          | 1305          |
| 143       | 4      | 3100       | 4281      | gb L42943       | Staphylococcus aureus (clone KIN50) phosphoenolpyruvate carboxykinase (pckA) gene, complete cds   | 100           | 1170          | 1182          |
| 143       | 5      | 4254       | 4718      | gb U51131       | Staphylococcus aureus phosphoenolpyruvate carboxykinase (pckA) gene, complete cds   | 100           | 449           | 465           |
| 143       | 9      | 6977       | 7261      | gb U51132       | Staphylococcus aureus o-succinylbenzoic acid CoA ligase (mene), and o-succinylbenzoic acid synthetase (mene) genes, complete cds        | 100           | 75            | 285           |
| 143       | 10     | 9464       | 8361      | gb U51132       | Staphylococcus aureus o-succinylbenzoic acid CoA ligase (mene), and o-succinylbenzoic acid synthetase (mene) genes, complete cds        | 100           | 1104          | 1104          |
| 143       | 11     | 11232      | 9748      | gb U51132       | Staphylococcus aureus o-succinylbenzoic acid CoA ligase (mene), and o-succinylbenzoic acid synthetase (mene) genes, complete cds        | 100           | 1485          | 1485          |
| 143       | 12     | 10739      | 10320     | gb U51132       | Staphylococcus aureus o-succinylbenzoic acid CoA ligase (mene), and o-succinylbenzoic acid synthetase (mene) genes, complete cds        | 100           | 332           | 420           |
| 152       | 5      | 2454       | 3437      | emb X58434 SAPD | S. aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydroliipoamide acetyltransferase and dihydroliipoamide dehydrogenase | 99            | 305           | 984           |
| 152       | 6      | 3513       | 4820      | emb X58434 SAPD | S. aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydroliipoamide acetyltransferase and dihydroliipoamide dehydrogenase | 98            | 1308          | 1308          |
| 152       | 7      | 4818       | 6230      | emb X58434 SAPD | S. aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydroliipoamide acetyltransferase and dihydroliipoamide dehydrogenase | 99            | 1413          | 1413          |
| 153       | 1      | 387        | 1526      | gb S77055       | recF cluster: dnaA-replisome assembly protein...gyrB-DNA gyrase beta subunit [Staphylococcus aureus, Y8886, Genomic, 5 genes, 3573 nt]  | 99            | 1140          | 1140          |
| 153       | 2      | 1877       | 2152      | gb L77055       | recF cluster: dnaA-replisome assembly protein...gyrB-DNA gyrase beta subunit [Staphylococcus aureus, Y8886, Genomic, 5 genes, 3573 nt]  | 100           | 276           | 276           |

TABLE 1

S. aureus - Coding regions containing known sequences

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | percent ident | IISF nt length | ORF nt length |
|-----------|--------|------------|-----------|-----------------|---|---------------|----------------|---------------|
| 153       | 3      | 2143       | 2289      | gb J77055       | tec cluster: dnaA-replisome assembly protein...gyrB-DNA gyrase beta subunit (Staphylococcus aureus, YB886, Genomic, 5 genes, 3573 nt) | 99            | 113            | 147           |
| 154       | 10     | 10792      | 9314      | gb J06451       | Staphylococcus aureus proline permease homolog (putP) gene, complete cds  | 91            | 154            | 1479          |
| 154       | 11     | 9935       | 9615      | gb J06451       | Staphylococcus aureus proline permease homolog (putP) gene, complete cds  | 99            | 229            | 321           |
| 154       | 12     | 9943       | 10167     | gb J06451       | Staphylococcus aureus proline permease homolog (putP) gene, complete cds  | 94            | 123            | 225           |
| 154       | 13     | 10089      | 11501     | gb J06451       | Staphylococcus aureus proline permease homolog (putP) gene, complete cds  | 99            | 1326           | 1413          |
| 159       | 2      | 2195       | 1212      | dbj D28879 STAP | Staphylococcus aureus gene for penicillin-binding protein 1, complete cds   | 100           | 71             | 984           |
| 161       | 3      | 2596       | 2270      | gb J83994       | Staphylococcus aureus prolipoprotein signal peptidase (lsp) gene, complete cds  | 92            | 203            | 327           |
| 162       | 1      | 1406       | 705       | gb J21221       | Staphylococcus aureus hyaluronate lyase (hyla) gene, complete cds   | 100           | 702            | 702           |
| 163       | 4      | 1263       | 1772      | gb J19770       | Staphylococcus aureus pyrrolidone carboxyl peptidase (pcp) gene, complete cds   | 96            | 127            | 510           |
| 164       | 7      | 4774       | 9117      | dbj D85727 D857 | Staphylococcus aureus DNA for DNA polymerase III, complete cds  | 99            | 3470           | 4344          |
| 168       | 7      | 7448       | 6447      | gb J21636       | Staphylococcus aureus cnp-binding-factor 1 (cbf1) and ORF X genes, complete cds   | 100           | 1002           | 1002          |
| 168       | 8      | 9538       | 7961      | gb J21636       | Staphylococcus aureus cnp-binding-factor 1 (cbf1) and ORF X genes, complete cds   | 99            | 1158           | 1578          |
| 173       | 6      | 9240       | 7801      | gb J03479       | S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds                       | 100           | 1440           | 1440          |
| 173       | 7      | 11252      | 9522      | gb J03479       | S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds                       | 99            | 1731           | 1731          |
| 173       | 8      | 8285       | 8704      | gb J03479       | S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds                       | 100           | 420            | 420           |
| 173       | 9      | 10168      | 9839      | gb J03479       | S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds                       | 100           | 330            | 330           |
| 173       | 10     | 11815      | 10829     | emb X14827 SALA | Staphylococcus aureus lacC and lacD genes   | 100           | 987            | 987           |
| 173       | 11     | 12721      | 11774     | emb X14827 SALA | Staphylococcus aureus lacC and lacD genes   | 100           | 948            | 948           |
| 173       | 12     | 12838      | 12305     | gb J64724       | S.aureus tagatase 6-phosphate isomerase gene, complete cds  | 100           | 534            | 534           |
| 173       | 13     | 13243      | 13773     | gb J32103       | Staphylococcus aureus lac repressor (lacI) gene, complete cds and lacA repressor (lacA), partial cds                                  | 100           | 471            | 471           |
| 173       | 14     | 14633      | 13866     | gb J32103       | Staphylococcus aureus lac repressor (lacI) gene, complete cds and lacA repressor (lacA), partial cds                                  | 100           | 768            | 768           |

TABLE 1

S. aureus - Coding regions containing known sequences

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession   | Match gene name  | Percent ident | HSP nt length | ORF nt length |
|-----------|--------|------------|-----------|-------------------|--|---------------|---------------|---------------|
| 178       | 1      | 2          | 655       | [gb U52961]       | Staphylococcus aureus holin-like protein LrgA (lrgA) and LrgB (lrgB) genes, complete cds   | 100           | 115           | 654           |
| 178       | 2      | 2201       | 1482      | [gb U52961]       | Staphylococcus aureus holin-like protein LrgA (lrgA) and LrgB (lrgB) genes, complete cds   | 100           | 720           | 720           |
| 178       | 3      | 2161       | 1909      | [gb U52961]       | Staphylococcus aureus holin-like protein LrgA (lrgA) and LrgB (lrgB) genes, complete cds   | 100           | 453           | 453           |
| 178       | 4      | 1551       | 1853      | [gb U52961]       | Staphylococcus aureus holin-like protein LrgA (lrgA) and LrgB (lrgB) genes, complete cds   | 100           | 303           | 303           |
| 178       | 5      | 3541       | 2777      | [gb U42945]       | Staphylococcus aureus lysS and lytX genes, complete cds  | 99            | 765           | 765           |
| 178       | 6      | 3294       | 3025      | [gb U42945]       | Staphylococcus aureus lysS and lytX genes, complete cds  | 99            | 270           | 270           |
| 181       | 1      | 1114       | 590       | [gb U61177]       | S. aureus sigma factor (pleC) gene, complete cds   | 99            | 499           | 525           |
| 182       | 1      | 3          | 341       | [emb X61307 SASP] | Staphylococcus aureus spa gene for protein A   | 98            | 277           | 339           |
| 182       | 2      | 690        | 2312      | [gb U01786]       | S. aureus spa gene coding for protein A, complete cds  | 97            | 1332          | 1623          |
| 182       | 3      | 5861       | 4251      | [emb X61307 SASP] | Staphylococcus aureus spa gene for protein A   | 99            | 119           | 1611          |
| 185       | 1      | 3          | 824       | [gb U31979]       | Staphylococcus aureus chorismate synthase (aroC) and nucleoside diphosphate kinase (ndk) genes, complete cds, dehydrogenase synthase (aroB) and geranylgeranyl pyrophosphate synthetase homolog (gerCC) genes, partial cds | 90            | 132           | 822           |
| 191       | 3      | 841        | 2760      | [emb X17679 SACO] | Staphylococcus aureus coa gene for coagulase   | 99            | 1920          | 1920          |
| 191       | 4      | 2967       | 3143      | [emb X16457 SAST] | Staphylococcus aureus gene for staphylocoagulase   | 99            | 177           | 177           |
| 191       | 5      | 5768       | 4566      | [emb X16457 SAST] | Staphylococcus aureus gene for staphylocoagulase   | 99            | 250           | 1203          |
| 196       | 1      | 1741       | 872       | [gb U36472]       | Staphylococcus aureus lysyl-tRNA synthetase gene, complete cds, transfer RNA (tRNA) genes, 5S ribosomal RNA (5S rRNA) gene, 16S ribosomal RNA (16S rRNA) gene, 23S ribosomal RNA (23S rRNA) gene                           | 99            | 870           | 870           |
| 198       | 3      | 1688       | 2011      | [emb X3205 SAPT]  | S. aureus ptbH and ptbI genes  | 99            | 324           | 324           |
| 198       | 4      | 2005       | 2310      | [emb X3205 SAPT]  | S. aureus ptbH and ptbI genes  | 97            | 304           | 306           |
| 202       | 1      | 163        | 1305      | [emb X97985 SA12] | S. aureus orfA 1.2.3 & 4   | 99            | 1143          | 1143          |
| 202       | 2      | 1303       | 2175      | [emb X73889 SAP1] | S. aureus genes P1 and P2  | 94            | 444           | 873           |
| 210       | 1      | 3114       | 1558      | [dbj D17366 STAA] | Staphylococcus aureus atl gene for autolysin, complete cds and other ORFs  | 99            | 1552          | 1557          |
| 210       | 2      | 2939       | 2232      | [gb U41499]       | Staphylococcus aureus ORF1, partial cds, ORF2, ORF3, autolysin (atl) genes, complete cds   | 99            | 684           | 708           |
| 214       | 11     | 7439       | 7770      | [dbj D86240 D862] | Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds   | 96            | 357           | 342           |

TABLE 1

S. aureus - Coding regions containing known sequences

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession   | Match gene name  | Percent ident | HSP nt length | ORF nt length |
|-----------|--------|------------|-----------|-------------------|--|---------------|---------------|---------------|
| 216       | 3      | 398        | 1318      | [emb]X72700[SAPV] | S. aureus genes for S and F components of Pantone-Valentine leucocidins  | 88            | 265           | 921           |
| 219       | 2      | 1810       | 1073      | [dbj]D10690[STAN] | Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds  | 100           | 60            | 738           |
| 219       | 3      | 2979       | 2035      | [dbj]D10690[STAN] | Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds  | 99            | 945           | 945           |
| 219       | 4      | 4359       | 3196      | [dbj]D10690[STAN] | Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds  | 99            | 1164          | 1164          |
| 219       | 5      | 7044       | 4176      | [dbj]D10690[STAN] | Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds  | 98            | 1869          | 1869          |
| 219       | 6      | 6557       | 5883      | [dbj]D10690[STAN] | Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds  | 99            | 675           | 675           |
| 219       | 7      | 6801       | 6334      | [dbj]D10690[STAN] | Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds  | 98            | 468           | 468           |
| 221       | 8      | 10816      | 10034     | [gb]U19298[ ]     | Staphylococcus aureus phosphatidylinositol-specific phospholipase C (plc)  | 91            | 67            | 783           |
| 223       | 1      | 2855       | 1506      | [gb]U73374[ ]     | Staphylococcus aureus type 8 capsule genes, cap8a, cap8b, cap8c, cap8d, cap8e, cap8f, cap8g, cap8h, cap8i, cap8j, cap8k, cap8l, cap8m, cap8n, cap8o, cap8p, complete cds | 99            | 102           | 1350          |
| 234       | 1      | 2          | 1337      | [emb]X97985[SA12] | S. aureus orfs 1.2.3 & 4   | 100           | 176           | 1350          |
| 234       | 2      | 1694       | 2485      | [emb]X97985[SA12] | S. aureus orfs 1.2.3 & 4   | 100           | 792           | 792           |
| 234       | 3      | 2648       | 3148      | [emb]X97985[SA12] | S. aureus orfs 1.2.3 & 4   | 99            | 501           | 501           |
| 234       | 4      | 3120       | 4604      | [emb]X97985[SA12] | S. aureus orfs 1.2.3 & 4   | 99            | 1305          | 1485          |
| 236       | 6      | 3826       | 5322      | [gb]U48826[ ]     | Staphylococcus aureus elastin binding protein (ebp) gene, complete cds   | 96            | 648           | 1497          |
| 248       | 1      | 2          | 403       | [emb]X62288[SAPV] | S. aureus DNA for penicillin-binding protein 2   | 100           | 103           | 402           |
| 248       | 2      | 388        | 852       | [gb]U35426[ ]     | Staphylococcus aureus penicillin-binding protein 2 (pbp2) gene, complete cds   | 99            | 465           | 465           |
| 253       | 2      | 1539       | 1093      | [gb]U46541[ ]     | Staphylococcus aureus sara gene, complete cds  | 96            | 447           | 447           |
| 254       | 2      | 150        | 1835      | [gb]U57060[ ]     | Staphylococcus aureus sda gene, complete cds   | 94            | 142           | 1686          |
| 254       | 3      | 1873       | 2728      | [gb]U57060[ ]     | Staphylococcus aureus sda gene, complete cds   | 99            | 756           | 756           |
| 260       | 1      | 2          | 1900      | [gb]M90693[ ]     | Staphylococcus aureus glycerol ester hydrolase (lip) gene, complete cds  | 99            | 1213          | 1899          |
| 265       | 1      | 1          | 942       | [dbj]D11131[STAS] | Staphylococcus aureus gene for a participant in homogeneous expression of high-level methicillin resistance, complete cds  | 99            | 941           | 942           |

TABLE I

S. aureus - Coding regions containing known sequences

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession  | match gene name   | percent ident | HSP nt length | ORF nt length |
|-----------|--------|------------|-----------|------------------|---|---------------|---------------|---------------|
| 265       | 2      | 688        | 476       | [dbj 021131 STAS | Staphylococcus aureus gene for a participant in homogeneous expression of high-level methicillin resistance, complete cds | 99            | 213           | 213           |
| 265       | 3      | 2418       | 1765      | [dbj 021131 STAS | Staphylococcus aureus gene for a participant in homogeneous expression of high-level methicillin resistance, complete cds | 98            | 65            | 654           |
| 266       | 1      | 2          | 1018      | [dbj 014711 STAM | Staphylococcus aureus HSP10 and HSP60 genes   | 98            | 743           | 1017          |
| 282       | 1      | 1          | 525       | [gb 572488]      | hmb-porphobilinogen synthase [Staphylococcus aureus, SAI959, Genomic, 1087 nt]  | 100           | 110           | 525           |
| 282       | 2      | 516        | 1502      | [gb 572488]      | hmb-porphobilinogen synthase [Staphylococcus aureus, SAI959, Genomic, 1087 nt]  | 100           | 952           | 987           |
| 284       | 1      | 3          | 170       | [gb M63176]      | Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds                                    | 98            | 84            | 168           |
| 284       | 2      | 282        | 1034      | [gb M63176]      | Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds                                    | 100           | 712           | 753           |
| 284       | 3      | 1028       | 2026      | [gb M63176]      | Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds                                    | 99            | 979           | 999           |
| 284       | 4      | 1990       | 2202      | [gb M63176]      | Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds                                    | 98            | 187           | 213           |
| 289       | 1      | 1536       | 1991      | [gb M24701]      | S. aureus SauIAI-restriction-enzyme and SauIAI-modification-enzyme genes, complete cds                                    | 99            | 338           | 456           |
| 303       | 1      | 2          | 868       | [gb L01055]      | Staphylococcus aureus gamma-hemolysin components A, B and C (hlgA, hlgB, hlgC) genes, complete cds                        | 99            | 867           | 867           |
| 303       | 2      | 1409       | 2343      | [gb L01055]      | Staphylococcus aureus gamma-hemolysin components A, B and C (hlgA, hlgB, hlgC) genes, complete cds                        | 100           | 975           | 975           |
| 303       | 3      | 2367       | 3161      | [gb L01055]      | Staphylococcus aureus gamma-hemolysin components A, B and C (hlgA, hlgB, hlgC) genes, complete cds                        | 99            | 793           | 795           |
| 305       | 1      | 2707       | 1355      | [dbj D17366 STAA | Staphylococcus aureus atl gene for autolysin, complete cds and other ORFs   | 99            | 1343          | 1353          |
| 311       | 1      | 2628       | 1335      | [gb L42945]      | Staphylococcus aureus lytC and lytR genes, complete cds   | 98            | 1314          | 1314          |
| 312       | 6      | 7019       | 7870      | [gb L14017]      | Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds                            | 74            | 351           | 852           |
| 323       | 1      | 1998       | 1003      | [gb U031175]     | Staphylococcus aureus D-specific D-2-hydroxyacid dehydrogenase (adh) gene, complete cds                                   | 98            | 996           | 996           |
| 326       | 1      | 1          | 237       | [emb V00356 SASP | Staphylococcus aureus V8 serine protease gene   | 100           | 108           | 237           |
| 338       | 1      | 687        | 388       | [emb X64389 SALE | S. aureus leuF-P83 gene for F component of leucocidin R   | 98            | 259           | 300           |
| 338       | 2      | 1828       | 1088      | [emb X64389 SALE | S. aureus leuF-P81 gene for F component of leucocidin R   | 97            | 137           | 741           |

TABLE 1

S. aureus - Coding regions containing known sequences

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name  | percent ident | ORF nt length |
|-----------|--------|------------|-----------|-----------------|--|---------------|---------------|
| 342       | 2      | 579        | 1754      | gb U06462       | Staphylococcus aureus SA4 Fts2 (fts2) gene, complete cds   | 100           | 1176          |
| 344       | 2      | 517        | 1248      | emb V01281 SAMU | S. aureus mRNA for nuclease  | 98            | 732           |
| 349       | 1      | 457        | 230       | gb W20393       | S. aureus bacteriophage phi-11 attachment site (attB)  | 96            | 172           |
| 353       | 1      | 1016       | 516       | gb H03994       | Staphylococcus aureus prolipoprotein signal peptidase (lisp) gene, complete cds                                      | 100           | 187           |
| 353       | 2      | 1582       | 1046      | gb H03994       | Staphylococcus aureus prolipoprotein signal peptidase (lisp) gene, complete cds                                      | 99            | 537           |
| 356       | 1      | 3          | 674       | gb U20503       | Staphylococcus aureus HMC class II analog gene, complete cds   | 75            | 671           |
| 361       | 1      | 1          | 903       | gb U19298       | Staphylococcus aureus phosphatidylinositol-specific phospholipase C (plc) gene, complete cds                         | 98            | 747           |
| 361       | 2      | 1103       | 1507      | gb U19298       | Staphylococcus aureus phosphatidylinositol-specific phospholipase C (plc) gene, complete cds                         | 97            | 68            |
| 373       | 1      | 3          | 1148      | emb X62288 SAPE | S. aureus DNA for penicillin-binding protein 2   | 99            | 1146          |
| 389       | 1      | 1904       | 1248      | emb X62282 SAT5 | S. aureus target site DNA for IS431 insertion  | 97            | 349           |
| 400       | 1      | 1          | 540       | emb X61716 SAIL | S. aureus hlb gene encoding sphingomyelinase   | 99            | 389           |
| 400       | 2      | 1693       | 1187      | emb X13404 SAHL | Staphylococcus aureus hlb gene for beta-hemolysin  | 99            | 178           |
| 408       | 1      | 1810       | 1049      | gb S76213       | esp23-alkaline shock protein 23 (methicillin resistant) [Staphylococcus aureus, 912, Genomic, 1360 nt]               | 99            | 163           |
| 418       | 1      | 2          | 217       | gb A1489        | Staphylococcus aureus ORF1, partial cds, ORF2, ORF3, autolysin (atl) genes, complete cds                             | 100           | 216           |
| 418       | 2      | 854        | 639       | dbj D17365 STAA | Staphylococcus aureus atl gene for autolysin, complete cds and other ORFs  | 100           | 188           |
| 431       | 2      | 1262       | 2509      | gb U43098       | Transposon Tn5404 and insertion sequences IS1181 and IS1182 (from Staphylococcus aureus) DNA                         | 99            | 1248          |
| 422       | 1      | 2          | 325       | gb K02985       | S. aureus (strain RM450) transposon Tn554 insertion site   | 96            | 200           |
| 427       | 1      | 865        | 434       | dbj D28879 STAP | Staphylococcus aureus gene for penicillin-binding protein 1, complete cds  | 100           | 432           |
| 427       | 2      | 1829       | 1122      | dbj D28879 STAP | Staphylococcus aureus gene for penicillin-binding protein 1, complete cds  | 100           | 151           |
| 435       | 1      | 2          | 808       | dbj D86240 D862 | Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds         | 100           | 556           |
| 435       | 2      | 832        | 999       | dbj D86240 D862 | Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds         | 100           | 134           |
| 436       | 1      | 1341       | 685       | emb X17688 SAPE | S. aureus factor essential for expression of methicillin resistance (femA) gene, complete cds, and trpA gene, 3' end | 97            | 657           |

TABLE 1

S. aureus - Coding regions containing known sequences

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name  | Percent ident | HSP nt length | ORF nt length |
|-----------|--------|------------|-----------|-----------------|--|---------------|---------------|---------------|
| 436       | 2      | 2403       | 1657      | emb N17688 SAFE | S. aureus factor essential for expression of methicillin resistance (fema) gene, complete cds, and trpA gene, 3' end                                 | 100           | 294           | 747           |
| 442       | 1      | 347        | 1300      | emb X72700 SAPV | S. aureus genes for S and F components of Pantan-Valentine leucocidins   | 84            | 204           | 954           |
| 445       | 2      | 3906       | 2178      | gb L01055       | Staphylococcus aureus gamma-hemolysin components A, B and C (hlgA, hlgB, hlgC) genes, complete cds   | 98            | 187           | 273           |
| 447       | 1      | 167        | 1078      | gb U19770       | Staphylococcus aureus pyrrolidone carboxyl peptidase (pcp) gene, complete cds  | 100           | 514           | 912           |
| 447       | 2      | 1176       | 1784      | gb U19770       | Staphylococcus aureus pyrrolidone carboxyl peptidase (pcp) gene, complete cds  | 96            | 597           | 609           |
| 454       | 3      | 7309       | 4319      | emb Z18652 SACF | S. aureus gene for clumping factor   | 75            | 653           | 2991          |
| 472       | 4      | 7896       | 5479      | gb L25288       | Staphylococcus aureus gyrase-like protein alpha and beta subunit (gria and griB) genes, complete cds   | 99            | 2418          | 2418          |
| 472       | 5      | 8120       | 6792      | gb L25288       | Staphylococcus aureus gyrase-like protein alpha and beta subunit (gria and griB) genes, complete cds   | 99            | 1328          | 1329          |
| 475       | 2      | 566        | 889       | emb X52543 SAAG | S. aureus agrA, agrB and hid genes   | 100           | 76            | 324           |
| 481       | 4      | 1922       | 1560      | emb X64172 SARP | S. aureus rplL, rplM, rplN and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains | 100           | 250           | 363           |
| 481       | 5      | 1244       | 1534      | emb X64172 SARP | S. aureus rplL, rplM, rplN and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains | 100           | 224           | 291           |
| 487       | 2      | 1388       | 1188      | gb H83994       | Staphylococcus aureus prolipoprotein signal peptidase (lapi) gene, complete cds  | 98            | 72            | 201           |
| 489       | 1      | 2737       | 1370      | gb U21221       | Staphylococcus aureus hyaluronate lyase (hyaA) gene, complete cds  | 99            | 1368          | 1368          |
| 503       | 2      | 1135       | 653       | gb H83994       | Staphylococcus aureus prolipoprotein signal peptidase (lapi) gene, complete cds  | 100           | 108           | 483           |
| 511       | 3      | 1613       | 2242      | gb L14017       | Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds   | 84            | 323           | 630           |
| 511       | 4      | 3122       | 2700      | gb S76213       | asp23-alkaline shock protein 23 (methicillin resistant) [Staphylococcus aureus, 912, Genomic, 1360 nt]   | 96            | 423           | 423           |
| 520       | 2      | 758        | 1297      | emb X72014 SAFI | S. aureus fib gene for fibrinogen-binding protein  | 99            | 540           | 540           |
| 520       | 3      | 1436       | 1801      | emb X72013 SAFI | S. aureus fib gene for fibrinogen-binding protein  | 99            | 221           | 366           |
| 526       | 1      | 2150       | 1092      | gb U17366 STAA  | Staphylococcus aureus atl gene for autolysin, complete cds and other ORFs  | 99            | 641           | 1059          |

TABLE 1

S. aureus - Coding regions containing known sequences

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name   | Percent ident | HSP nt length | ORF nt length |
|-----------|--------|------------|-----------|-----------------|---|---------------|---------------|---------------|
| 528       | 2      | 58         | 963       | gb U19300       | Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank   | 99            | 260           | 906           |
| 528       | 3      | 1098       | 2870      | gb U19300       | Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank   | 99            | 866           | 1773          |
| 530       | 1      | 3          | 434       | gb U19379       | Staphylococcus aureus chorismate synthase (aroC) and nucleoside diphosphate kinase (ndk) genes, complete cds, dehydroquininate synthase (aroB) and geranylgeranyl pyrophosphate synthetase homolog (gerCC) genes, partial cds | 99            | 432           | 432           |
| 530       | 2      | 1211       | 2395      | gb U19379       | Staphylococcus aureus chorismate synthase (aroC) and nucleoside diphosphate kinase (ndk) genes, complete cds, dehydroquininate synthase (aroB) and geranylgeranyl pyrophosphate synthetase homolog (gerCC) genes, partial cds | 91            | 1185          | 1185          |
| 530       | 3      | 2409       | 2801      | gb U19379       | Staphylococcus aureus chorismate synthase (aroC) and nucleoside diphosphate kinase (ndk) genes, complete cds, dehydroquininate synthase (aroB) and geranylgeranyl pyrophosphate synthetase homolog (gerCC) genes, partial cds | 88            | 181           | 393           |
| 530       | 4      | 2690       | 3484      | gb U05004       | Staphylococcus aureus dehydroquininate synthase (aroB) gene, 3' end cds; 3-phosphoshikimate-1-carboxyvinyltransferase (aroA) gene, complete cds; ORF3, complete cds   | 100           | 75            | 795           |
| 530       | 5      | 3482       | 4792      | gb U05004       | Staphylococcus aureus dehydroquininate synthase (aroB) gene, 3' end cds; 3-phosphoshikimate-1-carboxyvinyltransferase (aroA) gene, complete cds; ORF3, complete cds   | 99            | 905           | 1311          |
| 530       | 6      | 4790       | 5380      | gb U05004       | Staphylococcus aureus dehydroquininate synthase (aroB) gene, 3' end cds; 3-phosphoshikimate-1-carboxyvinyltransferase (aroA) gene, complete cds; ORF3, complete cds   | 100           | 196           | 591           |
| 539       | 1      | 3          | 338       | emb X76490 SACL | S. aureus (b270) glnA and glnR genes  | 99            | 336           | 336           |
| 539       | 2      | 336        | 527       | emb X76490 SACL | S. aureus (b270) glnA and glnR genes  | 100           | 189           | 192           |
| 554       | 1      | 727        | 365       | gb U73374       | Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds  | 100           | 54            | 363           |
| 554       | 2      | 2175       | 1252      | gb U73374       | Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds  | 99            | 918           | 924           |
| 554       | 3      | 1574       | 1374      | gb U73374       | Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds  | 96            | 122           | 201           |
| 584       | 2      | 1019       | 705       | gb U21221       | Staphylococcus aureus hyaluronate lyase (hlyA) gene, complete cds   | 99            | 306           | 315           |
| 587       | 3      | 1475       | 4286      | emb Z18852 SACF | S. aureus gene for clumping factor  | 98            | 2588          | 2814          |
| 598       | 1      | 3881       | 1953      | dif D28875 STAP | Staphylococcus aureus gene for penicillin-binding protein 1, complete cds   | 99            | 1873          | 1929          |

TABLE 1

S. aureus - Coding regions containing known sequences

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name  | Percent ident | ORF nt length |
|-----------|--------|------------|-----------|-----------------|--|---------------|---------------|
| 605       | 1      | 2          | 745       | dbj D86240 D862 | Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltU genes, complete cds               | 98            | 744           |
| 609       | 1      | 1628       | 816       | emb X76490 SAGU | S. aureus (bb270) glrA and glrR genes  | 100           | 813           |
| 614       | 1      | 1280       | 642       | gb J32103       | Staphylococcus aureus lac repressor (lacR) gene, complete cds and lacA repressor (lacA), partial cds                       | 99            | 639           |
| 626       | 1      | 2508       | 1255      | gb H63176       | Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds                                     | 100           | 1254          |
| 626       | 2      | 3315       | 2284      | gb H63176       | Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds                                     | 99            | 1032          |
| 629       | 1      | 1995       | 1001      | emb X17688 SAFE | S. aureus factor essential for expression of methicillin resistance (femA) gene, complete cds, and trpA gene, 3' end       | 99            | 999           |
| 629       | 2      | 1407       | 1195      | emb X17688 SAFE | S. aureus factor essential for expression of methicillin resistance (femA) gene, complete cds, and trpA gene, 3' end       | 98            | 213           |
| 631       | 2      | 5126       | 3228      | emb Z18852 SACF | S. aureus gene for clumping factor   | 82            | 1899          |
| 632       | 1      | 3          | 551       | emb Z10568 SAST | S. aureus (RN4220) genes for potential ABC transporter and potential membrane spanning protein                             | 99            | 549           |
| 632       | 2      | 529        | 1323      | emb Z10568 SAST | S. aureus (RN4220) genes for potential ABC transporter and potential membrane spanning protein                             | 99            | 795           |
| 651       | 1      | 1909       | 1070      | gb L19100       | Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank          | 99            | 840           |
| 657       | 2      | 1800       | 1105      | gb L14017       | Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds                             | 84            | 696           |
| 662       | 1      | 908        | 456       | emb X13404 SAHL | Staphylococcus aureus hlb gene for beta-hemolysin  | 100           | 453           |
| 662       | 2      | 230        | 475       | emb X13404 SAHL | Staphylococcus aureus hlb gene for beta-hemolysin  | 100           | 246           |
| 662       | 3      | 746        | 1399      | emb X13404 SAHL | Staphylococcus aureus hlb gene for beta-hemolysin  | 99            | 654           |
| 682       | 1      | 956        | 480       | gb H63177       | S. aureus sigma factor (plac) gene, complete cds   | 100           | 477           |
| 685       | 1      | 1182       | 592       | gb J65000       | Staphylococcus aureus type-I signal peptidase SpaA (spaA) gene, and type-I signal peptidase SpSB (spSB) gene, complete cds | 98            | 591           |
| 685       | 2      | 1716       | 1153      | gb J65000       | Staphylococcus aureus type-I signal peptidase SpaA (spaA) gene, and type-I signal peptidase SpSB (spSB) gene, complete cds | 96            | 564           |
| 697       | 1      | 3          | 527       | gb H63177       | S. aureus sigma factor (plac) gene, complete cds   | 100           | 525           |
| 697       | 2      | 485        | 784       | gb H63177       | S. aureus sigma factor (plac) gene, complete cds   | 97            | 300           |

TABLE 1

S. aureus - Coding regions containing known sequences

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession  | Match gene name  | Percent ident | HSP nt length | ORF nt length |
|-----------|--------|------------|-----------|------------------|--|---------------|---------------|---------------|
| 710       | 1      | 15         | 503       | [dbj]DR6240 DR62 | Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds   | 99            | 217           | 489           |
| 733       | 1      | 26         | 205       | [gb]M802521      | Staphylococcus aureus norA199 gene (which mediates active efflux of fluoroquinolones), complete cds  | 97            | 140           | 180           |
| 741       | 1      | 1736       | 1197      | [dbj]DR1951 STAL | Staphylococcus aureus DNA for LUKK component, LUKF-PV like component, complete cds   | 81            | 522           | 540           |
| 752       | 1      | 1          | 636       | [emb]Y00356 SASP | Staphylococcus aureus V8 serine protease gene  | 99            | 618           | 636           |
| 752       | 2      | 588        | 956       | [emb]Y00356 SASP | Staphylococcus aureus V8 serine protease gene  | 99            | 340           | 369           |
| 756       | 1      | 1308       | 709       | [emb]X01645 SAT0 | Staphylococcus aureus (Wood 46) gene for alpha-toxin   | 98            | 567           | 600           |
| 777       | 1      | 1582       | 950       | [emb]Z49245 SA42 | S. aureus partial sod gene for superoxide dismutase  | 99            | 429           | 633           |
| 780       | 1      | 1111       | 557       | [gb]U205031      | Staphylococcus aureus HMC class II analog gene, complete cds   | 86            | 550           | 555           |
| 784       | 1      | 73         | 687       | [gb]U633291      | Staphylococcus aureus novel antigen gene, complete cds   | 99            | 588           | 615           |
| 797       | 1      | 182        | 544       | [dbj]D14711 STAH | Staphylococcus aureus HSP10 and HSP60 genes  | 98            | 363           | 363           |
| 798       | 1      | 532        | 302       | [emb]X58434 SNPD | S. aureus pddH, pddC and pddD genes for pyruvate decarboxylase, dihydrolysoamide acetyltransferase and dihydrolysoamide dehydrogenase                | 95            | 156           | 231           |
| 823       | 1      | 3          | 467       | [gb]S770551      | recF cluster: dnaArep1000 assembly protein...gyrB-DNA gyrase beta subunit [Staphylococcus aureus, Y8886, Genomic, 5 genes, 3573 nt]                  | 99            | 156           | 465           |
| 848       | 1      | 348        | 175       | [gb]L252881      | Staphylococcus aureus gyrase-like protein alpha and beta subunit (gr1A and gr1B) genes, complete cds   | 99            | 174           | 174           |
| 848       | 2      | 476        | 318       | [gb]L252881      | Staphylococcus aureus gyrase-like protein alpha and beta subunit (gr1A and gr1B) genes, complete cds   | 100           | 131           | 159           |
| 866       | 1      | 792        | 397       | [emb]X64172 SARP | S. aureus rplL, rplM, rplN and rplP genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta L beta' chains | 99            | 395           | 396           |
| 883       | 1      | 1          | 285       | [dbj]D90119 STAM | S. aureus nraA gene  | 99            | 131           | 285           |
| 884       | 1      | 606        | 334       | [emb]X52543 SAG  | S. aureus agrA, agrB and hld genes   | 98            | 265           | 273           |
| 884       | 2      | 716        | 522       | [emb]X52543 SAG  | S. aureus agrA, agrB and hld genes   | 100           | 195           | 195           |
| 912       | 2      | 517        | 681       | [emb]Z30588 SAST | S. aureus (RM4220) genes for potential ABC transporter and potential membrane spanning protein   | 99            | 163           | 165           |
| 917       | 1      | 2          | 265       | [gb]M647241      | S. aureus tagatase 6-phosphate isomerase gene, complete cds  | 99            | 247           | 264           |
| 917       | 2      | 236        | 396       | [gb]M647241      | S. aureus tagatase 6-phosphate isomerase gene, complete cds  | 95            | 147           | 159           |
| 918       | 1      | 2426       | 1215      | [emb]X93205 SAPT | S. aureus ptmH and ptmI genes  | 99            | 1212          | 1212          |

TABLE I

S. aureus - Coding regions containing known sequences

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession  | Match gene name   | percent ident | ORF nt length |
|-----------|--------|------------|-----------|------------------|---|---------------|---------------|
| 967       | 1      | 1          | 411       | [dbj 090119 STAM | S. aureus nraA gene   | 97            | 395           |
| 991       | 1      | 672        | 337       | [emb X52543 SAG  | S. aureus agrA, agrB and hid genes  | 99            | 336           |
| 1000      | 1      | 1117       | 845       | [gb L14017]      | Staphylococcus aureus methicillin-resistance protein (mecA) gene and unknown ORF, complete cds  | 78            | 190           |
| 1001      | 1      | 498        | 265       | [dbj D86240 D862 | Staphylococcus aureus gene for unknown function and dit operon dltA, dltB, dltC and dltD genes, complete cds                            | 99            | 234           |
| 1010      | 1      | 1          | 285       | [gb U21221]      | Staphylococcus aureus hyaluronate lyase (hysA) gene, complete cds   | 99            | 224           |
| 1046      | 1      | 656        | 330       | [emb X17200 SAPV | S. aureus genes for S and P components of Pantone-Valentine leukocidins   | 85            | 205           |
| 1040      | 1      | 480        | 286       | [emb X58434 SAPD | S. aureus pdhA, pdhC and pdhD genes for pyruvate decarboxylase, dihydroliipoamide acetyltransferase and dihydroliipoamide dehydrogenase | 99            | 180           |
| 1073      | 1      | 1176       | 589       | [gb K02985]      | S. aureus (strain RN450) transposon Tn534 insertion site  | 100           | 131           |
| 1079      | 1      | 3          | 230       | [dbj D86240 D862 | Staphylococcus aureus gene for unknown function and dit operon dltA, dltB, dltC and dltD genes, complete cds                            | 99            | 228           |
| 1079      | 2      | 218        | 484       | [dbj D86240 D862 | Staphylococcus aureus gene for unknown function and dit operon dltA, dltB, dltC and dltD genes, complete cds                            | 100           | 267           |
| 1079      | 3      | 460        | 645       | [dbj D86240 D862 | Staphylococcus aureus gene for unknown function and dit operon dltA, dltB, dltC and dltD genes, complete cds                            | 100           | 186           |
| 1092      | 1      | 289        | 146       | [emb X58434 SAPD | S. aureus pdhA, pdhC and pdhD genes for pyruvate decarboxylase, dihydroliipoamide acetyltransferase and dihydroliipoamide dehydrogenase | 98            | 124           |
| 1143      | 1      | 1          | 243       | [gb H63177]      | S. aureus alpmo (actor) (placI) gene, complete cds  | 99            | 243           |
| 1152      | 1      | 2          | 136       | [emb Z48003 SADW | S. aureus gene for DNA polymerase III   | 97            | 127           |
| 1189      | 1      | 720        | 361       | [gb S74031]      | norA-NorA (ISP794) [Staphylococcus aureus, NCTC 8325, Insertion, 1820 nt]   | 99            | 360           |
| 1190      | 1      | 2          | 283       | [gb H21854]      | S. aureus agr gene encoding an accessory gene regulator protein, complete cds   | 100           | 282           |
| 1190      | 2      | 1127       | 888       | [emb X52543 SAG  | S. aureus agrA, agrB and hid genes  | 100           | 240           |
| 1225      | 1      | 2          | 163       | [emb X17679 SACO | Staphylococcus aureus coa gene for coagulase  | 97            | 124           |
| 1243      | 1      | 2          | 529       | [dbj D86240 D862 | Staphylococcus aureus gene for unknown function and dit operon dltA, dltB, dltC and dltD genes, complete cds                            | 99            | 495           |
| 1244      | 1      | 1          | 210       | [gb S74031]      | norA-NorA (ISP794) [Staphylococcus aureus, NCTC 8325, Insertion, 1820 nt]   | 100           | 210           |
| 1301      | 1      | 41         | 472       | [emb X76490 ENGL | S. aureus (bb270) glpA and glpB genes   | 99            | 299           |

TABLE I

5. aureus - coding regions containing known sequences

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name  | Percent Ident | HSP nt length | ORF nt length |
|-----------|--------|------------|-----------|-----------------|--|---------------|---------------|---------------|
| 1315      | 1      | 18         | 336       | emb X64172 SARP | S. aureus rpl, orf202, rpoB genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chain  | 98            | 277           | 309           |
| 1519      | 1      | 2          | 175       | dbj D28879 STAP | Staphylococcus aureus gene for penicillin-binding protein 1, complete cds  | 98            | 139           | 174           |
| 1663      | 1      | 1346       | 675       | dbj D64240 D862 | Staphylococcus aureus gene for unknown function and dit operon dltA, dltB, dltC and dltD genes, complete cds   | 98            | 672           | 672           |
| 1797      | 1      | 644        | 324       | gb U73374       | Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds   | 99            | 321           | 321           |
| 1857      | 1      | 1          | 192       | gb H90336       | Staphylococcus aureus alpha-hemolysin gene, 3' end   | 98            | 192           | 192           |
| 1923      | 1      | 2          | 181       | emb X17688 SAFE | S. aureus factor essential for expression of methicillin resistance (fema) gene, complete cds, and trpA gene, 3' end   | 100           | 180           | 180           |
| 1957      | 1      | 2          | 346       | gb U60589       | Staphylococcus aureus novel antigen gene, complete cds   | 99            | 345           | 345           |
| 1988      | 1      | 1          | 402       | dbj D64240 D862 | Staphylococcus aureus gene for unknown function and dit operon dltA, dltB, dltC and dltD genes, complete cds   | 100           | 402           | 402           |
| 2100      | 1      | 414        | 208       | gb H63177       | S. aureus sigma factor (plac) gene, complete cds   | 99            | 207           | 207           |
| 2199      | 1      | 1          | 402       | gb U66664       | Staphylococcus aureus DNA fragment with class II promoter activity   | 99            | 131           | 402           |
| 2537      | 1      | 308        | 156       | emb X17688 SAFE | S. aureus factor essential for expression of methicillin resistance (fema) gene, complete cds, and trpA gene, 3' end   | 99            | 153           | 153           |
| 2891      | 1      | 2          | 400       | gb U25426       | Staphylococcus aureus penicillin-binding protein 2 (pbp2) gene, complete cds   | 99            | 399           | 399           |
| 2950      | 1      | 778        | 398       | dbj D30690 STAN | Staphylococcus aureus genes for ORF37, HSP20, HSP70, HSP40, ORF35, complete cds  | 100           | 358           | 381           |
| 2971      | 1      | 3          | 398       | gb U51132       | Staphylococcus aureus o-succinylbenzoic acid CoA ligase (menc), and o-succinylbenzoic acid synthetase (menc) genes, complete cds   | 97            | 272           | 396           |
| 2978      | 1      | 618        | 328       | gb U31979       | Staphylococcus aureus chorismate synthase (aroC) and nucleoside diphosphate kinase (ndk) genes, complete cds, dehydrogenase synthase (aroB) and geranylgeranyl pyrophosphate synthetase homolog (gerCC) genes, partial cds | 98            | 250           | 291           |
| 2985      | 1      | 832        | 464       | emb X17679 SACO | Staphylococcus aureus coa gene for coagulase   | 98            | 347           | 369           |
| 3006      | 1      | 2170       | 1784      | gb U11779       | Staphylococcus aureus methicillin-resistant ATCC 33952 clone RANV30 16S-23S rRNA spacer region   | 87            | 82            | 387           |
| 3008      | 1      | 474        | 238       | dbj D30690 STAN | Staphylococcus aureus genes for ORF37, HSP20, HSP70, HSP40, ORF35, complete cds  | 88            | 178           | 237           |
| 3008      | 2      | 451        | 281       | dbj U30690 STAN | Staphylococcus aureus genes for ORF37, HSP20, HSP70, HSP40, ORF35, complete cds  | 97            | 120           | 171           |

TABLE 1

S aureus - Coding regions containing known sequences

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name  | Percent ident | HSP length | ORF nt length |
|-----------|--------|------------|-----------|-----------------|--|---------------|------------|---------------|
| 3011      | 1      | 793        | 398       | emb X62992 SAPN | S.aureus fnbB gene for fibronectin binding protein B   | 93            | 72         | 396           |
| 3019      | 1      | 2          | 235       | gb J03479       | S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds  | 97            | 234        | 234           |
| 3023      | 1      | 81         | 233       | gb U06451       | Staphylococcus aureus proline permease homolog (putP) gene, complete cds   | 87            | 100        | 153           |
| 3029      | 1      | 90         | 287       | gb U51133       | Staphylococcus aureus phosphoenolpyruvate carboxylase (pckA) gene, complete cds  | 100           | 135        | 198           |
| 3039      | 1      | 18         | 164       | gb U51133       | Staphylococcus aureus phosphoenolpyruvate carboxylase (pckA) gene, complete cds  | 97            | 135        | 147           |
| 3039      | 2      | 70         | 327       | gb U51133       | Staphylococcus aureus phosphoenolpyruvate carboxylase (pckA) gene, complete cds  | 77            | 183        | 258           |
| 3056      | 1      | 3          | 215       | emb X64172 SARP | S.aureus rplL, orf202, rpoB11f and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains | 99            | 213        | 213           |
| 3059      | 1      | 1          | 261       | dbj D30690 STAN | Staphylococcus aureus genes for ORF7; HSP70; HSP40; ORF35, complete cds  | 98            | 234        | 261           |
| 3073      | 1      | 27         | 284       | gb U06451       | Staphylococcus aureus proline permease homolog (putP) gene, complete cds   | 99            | 229        | 258           |
| 3074      | 1      | 2          | 397       | emb X64172 SARP | S.aureus rplL, orf202, rpoB11f and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains | 96            | 350        | 396           |
| 3088      | 1      | 3          | 239       | dbj D6727 DR67  | Staphylococcus aureus DNA for DNA polymerase III, complete cds   | 95            | 215        | 237           |
| 3097      | 1      | 444        | 244       | emb Z48003 SAPN | S.aureus gene for DNA polymerase III   | 97            | 160        | 201           |
| 3102      | 1      | 307        | 155       | gb J03479       | S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds  | 97            | 142        | 153           |
| 3121      | 1      | 568        | 398       | emb X58434 SAPD | S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydroliposamide acetyltransferase and dihydroliposamide dehydrogenase                   | 100           | 88         | 171           |
| 3125      | 1      | 463        | 233       | emb X89233 SAMP | S.aureus DNA for rpoC gene   | 98            | 192        | 231           |
| 3133      | 1      | 2          | 175       | emb Z18852 SACP | S.aureus gene for clumping factor  | 96            | 154        | 174           |
| 3160      | 1      | 420        | 211       | dbj J10489 STAG | Staphylococcus aureus genes for DNA gyrase A and B, complete cds   | 89            | 197        | 210           |
| 3176      | 1      | 1          | 378       | emb X36434 SAPD | S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydroliposamide acetyltransferase and dihydroliposamide dehydrogenase                   | 96            | 91         | 378           |
| 3192      | 1      | 420        | 211       | gb J03479       | S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds  | 98            | 72         | 210           |
| 3210      | 1      | 3          | 143       | gb J76714       | Staphylococcus aureus peptidoglycan hydrolase gene, complete cds   | 96            | 141        | 141           |

TABLE I

S. aureus - Coding regions containing known sequences

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name   | Percent ident | ORF nt length |
|-----------|--------|------------|-----------|-----------------|---|---------------|---------------|
| 3332      | 3      | 2106       | 1282      | gb U14017       | Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds  | 71            | 257           |
| 353A      | 1      | 2          | 394       | emb X89233 SARP | S. aureus DNA for rpoC gene   | 99            | 390           |
| 3543      | 1      | 392        | 634       | gb U1530        | Staphylococcus aureus transfer RNA sequence with two RNAs   | 99            | 102           |
| 3555      | 1      | 637        | 320       | emb Z18852 SACF | S. aureus gene for clumping factor  | 99            | 307           |
| 3559      | 1      | 3          | 182       | emb X17679 SACO | Staphylococcus aureus coa gene for coagulase  | 100           | 141           |
| 3559      | 2      | 95         | 313       | emb X17679 SACO | Staphylococcus aureus coa gene for coagulase  | 98            | 174           |
| 3563      | 1      | 278        | 141       | gb U35733       | Staphylococcus aureus prolly-protein diacylglycerol transferase (lgt) gene, complete cds  | 100           | 79            |
| 3563      | 2      | 527        | 363       | gb U35733       | Staphylococcus aureus prolly-protein diacylglycerol transferase (lgt) gene, complete cds  | 98            | 162           |
| 3566      | 1      | 3          | 422       | emb X16457 SAST | Staphylococcus aureus gene for staphylocoagulase  | 98            | 175           |
| 3588      | 1      | 2          | 262       | gb U43098       | Transposon Tn5604 and insertion sequences IS1181 and IS1182 (from Staphylococcus aureus) DNA  | 99            | 253           |
| 3603      | 1      | 3          | 350       | gb J03479       | S. aureus enzyme III-lac (lacF), enzyme III-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds   | 99            | 345           |
| 3600      | 1      | 758        | 381       | emb Z18852 SACF | S. aureus gene for clumping factor  | 72            | 346           |
| 3602      | 1      | 788        | 396       | emb Z18852 SACF | S. aureus gene for clumping factor  | 98            | 319           |
| 3656      | 1      | 1013       | 528       | emb Z18852 SACF | S. aureus gene for clumping factor  | 84            | 403           |
| 3682      | 1      | 3          | 236       | emb X64172 SARP | S. aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains | 100           | 231           |
| 3682      | 2      | 224        | 415       | emb X64172 SARP | S. aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains | 100           | 112           |
| 3693      | 1      | 758        | 423       | emb X62992 SAPN | S. aureus fimb gene for fibronectin binding protein B   | 100           | 229           |
| 3702      | 1      | 593        | 354       | gb U11530       | Staphylococcus aureus transfer RNA sequence with two RNAs   | 94            | 81            |
| 3725      | 1      | 924        | 463       | emb Z18852 SACF | S. aureus gene for clumping factor  | 71            | 367           |
| 3761      | 1      | 809        | 450       | gb U14017       | Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds  | 85            | 333           |
| 3767      | 1      | 1          | 402       | emb X64172 SARP | S. aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains | 98            | 387           |

TABLE 1

aureus - Coding regions containing known sequences

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name   | Percent ident | ORF nt length |
|-----------|--------|------------|-----------|-----------------|---|---------------|---------------|
| 3775      | 1      | 2          | 286       | emb X64172 SARP | S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains        | 100           | 227           |
| 3786      | 1      | 456        | 229       | dbj D10489 STAG | Staphylococcus aureus genes for DNA gyrase A and B, complete cds  | 100           | 204           |
| 3786      | 2      | 532        | 366       | dbj D10489 STAG | Staphylococcus aureus genes for DNA gyrase A and B, complete cds  | 95            | 123           |
| 3798      | 1      | 3          | 251       | emb X17679 SACO | Staphylococcus aureus coa gene for coagulase  | 99            | 249           |
| 3813      | 1      | 793        | 398       | gb J04151       | S.aureus fibronectin-binding protein (fnbA) mRNA, complete cds  | 98            | 396           |
| 3819      | 1      | 184        | 402       | emb X68425 SA23 | S.aureus gene for 23S rRNA  | 99            | 161           |
| 3844      | 1      | 932        | 468       | gb U48826       | Staphylococcus aureus elastin binding protein (ebpS) gene, complete cds   | 87            | 204           |
| 3845      | 1      | 1          | 381       | emb X58434 SAPD | S.aureus pdhH, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase                              | 94            | 356           |
| 3856      | 1      | 798        | 400       | gb L14037       | Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds  | 76            | 192           |
| 3859      | 1      | 1049       | 573       | emb Z18852 SNCP | S.aureus gene for clumping factor   | 85            | 347           |
| 3871      | 1      | 650        | 327       | gb H76714       | Staphylococcus aureus peptidoglycan hydrolase gene, complete cds  | 100           | 299           |
| 3876      | 1      | 2          | 253       | dbj D10689 STAG | Staphylococcus aureus genes for DNA gyrase A and B, complete cds  | 100           | 217           |
| 3877      | 1      | 572        | 288       | gb J03479       | S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds   | 97            | 209           |
| 3878      | 1      | 1          | 237       | emb X58434 SAPD | S.aureus pdhH, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase                              | 96            | 155           |
| 3888      | 1      | 3          | 173       | emb X16457 SAST | Staphylococcus aureus gene for staphylocoagulase  | 98            | 171           |
| 3893      | 1      | 1          | 183       | emb X69233 SARP | S.aureus DNA for rpoC gene  | 100           | 170           |
| 3893      | 2      | 181        | 357       | emb X69233 SARP | S.aureus DNA for rpoC gene  | 98            | 79            |
| 3894      | 1      | 3          | 485       | emb X64172 SARP | S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains        | 99            | 450           |
| 3895      | 1      | 836        | 420       | gb J04151       | S.aureus fibronectin-binding protein (fnbA) mRNA, complete cds  | 99            | 411           |
| 3905      | 1      | 48         | 239       | gb L05004       | Staphylococcus aureus dehydroquinase synthase (aroB) gene, 3' end cds; 3-phosphoshikimate-1-carboxyvinyltransferase (aroC) gene, complete cds; ORF3, complete cds | 100           | 159           |
| 3905      | 2      | 188        | 400       | gb L05004       | Staphylococcus aureus dehydroquinase synthase (aroB) gene, 3' end cds; 3-phosphoshikimate-1-carboxyvinyltransferase (aroC) gene, complete cds; ORF3, complete cds | 97            | 86            |

TABLE 1

S. aureus - Coding regions containing known sequences

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name  | percent ident | HSP nt length | ORF nt length |
|-----------|--------|------------|-----------|-----------------|--|---------------|---------------|---------------|
| 3910      | 1      | 3          | 359       | emb X58434 SAPD | [S. aureus pdbB, pdhC and pddH genes for pyruvate decarboxylase, dihydroliipoamide acetyltransferase and dihydroliipoamide dehydrogenase                                 | 99            | 278           | 357           |
| 3915      | 1      | 1          | 330       | gb L14017       | Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds   | 75            | 175           | 330           |
| 3964      | 1      | 691        | 347       | emb Z48003 SADM | [S. aureus gene for DNA polymerase III   | 100           | 295           | 345           |
| 4007      | 1      | 199        | 390       | emb X16457 SAST | [Staphylococcus aureus gene for staphylocoagulase  | 98            | 163           | 192           |
| 4036      | 1      | 3          | 371       | dbj D10489 STAC | [Staphylococcus aureus genes for DNA gyrase A and B, complete cds  | 99            | 339           | 369           |
| 4046      | 1      | 692        | 348       | emb Z18852 SACF | [S. aureus gene for clumping factor  | 87            | 221           | 345           |
| 4060      | 1      | 1          | 375       | emb Z18852 SACF | [S. aureus gene for clumping factor  | 96            | 271           | 375           |
| 4061      | 1      | 860        | 432       | emb Z48003 SADM | [S. aureus gene for DNA polymerase III   | 99            | 429           | 429           |
| 4062      | 1      | 606        | 304       | gb L14017       | [Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds  | 75            | 198           | 303           |
| 4085      | 1      | 58         | 402       | gb U11766       | [Staphylococcus aureus methicillin-resistant ATCC 33932 clone RR442 16S-23S rRNA spacer region   | 98            | 127           | 345           |
| 4088      | 1      | 2          | 301       | gb L14058       | Transposon Tn5404 and insertion sequences IS1181 and IS1182 (from Staphylococcus aureus) DNA   | 99            | 227           | 300           |
| 4093      | 1      | 2          | 277       | emb X58434 SAPD | [S. aureus pddH, pdhC and pddH genes for pyruvate decarboxylase, dihydroliipoamide acetyltransferase and dihydroliipoamide dehydrogenase                                 | 99            | 276           | 276           |
| 4097      | 1      | 1          | 402       | emb Z18852 SACF | [S. aureus gene for clumping factor  | 74            | 307           | 402           |
| 4116      | 1      | 22         | 402       | gb L05004       | Staphylococcus aureus dehydroquinase synthase (aroH) gene, 3' end cds; 3-phosphohistidine-1-carboxyvinyltransferase (aroA) gene, complete cds; ORF3, complete cds        | 98            | 157           | 381           |
| 4125      | 1      | 240        | 401       | gb U73374       | Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds | 100           | 86            | 162           |
| 4149      | 1      | 35         | 247       | gb J04151       | [S. aureus fibronectin-binding protein (fba) mRNA, complete cds  | 99            | 200           | 213           |
| 4151      | 1      | 629        | 366       | gb L14017       | Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds   | 87            | 150           | 264           |
| 4154      | 1      | 754        | 398       | emb X64172 SARP | [S. aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains             | 99            | 297           | 357           |
| 4179      | 1      | 1          | 294       | emb X64172 SARP | [S. aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains             | 98            | 240           | 294           |

TABLE 1

S. aureus - Coding regions containing known sequences

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name  | percent ident | HSP nt length | ORF nt length |
|-----------|--------|------------|-----------|-----------------|--|---------------|---------------|---------------|
| 4203      | 1      | 1          | 255       | emb X8233 SARP  | S.aureus DNA for rpoC gene   | 99            | 239           | 255           |
| 4206      | 1      | 1          | 303       | emb Z18652 SACF | S.aureus gene for clumping factor  | 100           | 236           | 303           |
| 4206      | 2      | 195        | 344       | emb Z18652 SACF | S.aureus gene for clumping factor  | 95            | 65            | 150           |
| 4208      | 1      | 108        | 314       | emb X58434 SAPD | S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase                       | 89            | 76            | 207           |
| 4216      | 1      | 656        | 330       | emb X58434 SAPD | S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase                       | 98            | 326           | 327           |
| 4226      | 1      | 594        | 298       | gb L11530       | Staphylococcus aureus transfer RNA sequence with two rRNAs   | 97            | 132           | 297           |
| 4260      | 1      | 216        | 383       | gb J011784      | Staphylococcus aureus methicillin-resistant ATCC 33952 clone RMV40 16S-23S rRNA spacer region  | 83            | 141           | 168           |
| 4272      | 1      | 355        | 179       | emb Z48003 SADN | S.aureus gene for DNA polymerase III   | 100           | 164           | 177           |
| 4276      | 1      | 4          | 177       | emb X16457 SAST | Staphylococcus aureus gene for staphylocoagulase   | 99            | 150           | 174           |
| 4277      | 1      | 1          | 270       | emb X64172 SARP | S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains | 99            | 265           | 270           |
| 4282      | 1      | 691        | 377       | emb X64172 SARP | S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains | 98            | 282           | 315           |
| 4291      | 1      | 379        | 181       | emb X64172 SARP | S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains | 99            | 183           | 189           |
| 4295      | 1      | 3          | 329       | emb X16457 SAST | Staphylococcus aureus gene for staphylocoagulase   | 94            | 146           | 327           |
| 4313      | 1      | 435        | 280       | gb L11530       | Staphylococcus aureus transfer RNA sequence with two rRNAs   | 100           | 94            | 156           |
| 4315      | 1      | 3          | 185       | gb J03479       | S.aureus enzyme III-lac (lacP), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds  | 100           | 158           | 183           |
| 4315      | 2      | 101        | 310       | gb J03479       | S.aureus enzyme III-lac (lacP), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds  | 98            | 75            | 210           |
| 4327      | 1      | 1          | 294       | gb L143098      | Tnampson Tn5404 and insertion sequences IS1181 and IS1182 (from Staphylococcus aureus) DNA   | 98            | 294           | 294           |
| 4360      | 1      | 603        | 319       | gb U02910       | Staphylococcus aureus ATCC 25923 16S rRNA gene, partial sequence   | 100           | 116           | 285           |
| 4364      | 1      | 3          | 146       | emb X64172 SARP | S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains | 95            | 140           | 144           |
| 4388      | 1      | 167        | 310       | emb X62392 SAPN | S.aureus fnbB gene for fibronectin binding protein B   | 73            | 119           | 144           |

TABLE I

S. aureus - Coding regions containing known sequences

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name  | percent ident | ORF nt length |
|-----------|--------|------------|-----------|-----------------|--|---------------|---------------|
| 4401      | 1      | 2          | 313       | emb X62992 SAFN | S.aureus fnbB gene for fibronectin binding protein B   | 97            | 243           |
| 4421      | 1      | 36         | 281       | dbj D12572 STA2 | Staphylococcus aureus rna gene for 23S ribosomal RNA   | 100           | 112           |
| 4426      | 1      | 3          | 293       | emb Z18852 SACF | S.aureus gene for clumping factor  | 85            | 185           |
| 4428      | 1      | 493        | 248       | emb X64172 SARP | S.aureus rplL, rpf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains | 100           | 139           |
| 4462      | 1      | 2          | 271       | emb X64172 SARP | S.aureus rplL, rpf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains | 99            | 270           |
| 4466      | 1      | 1          | 240       | emb Z18852 SACF | S.aureus gene for clumping factor  | 99            | 231           |
| 4469      | 1      | 1          | 312       | gb J03479       | S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds  | 99            | 265           |
| 4485      | 1      | 3          | 263       | gb L43098       | Transposon Tn5404 and insertion sequences IS1181 and IS1182 (from Staphylococcus aureus) DNA   | 98            | 259           |
| 4492      | 1      | 74         | 400       | gb H66227       | Staphylococcus aureus DNA gyrase B subunit (gyrB) RecF homologue (racF) and DNA gyrase A subunit (gyrA) gene, complete cds                                 | 85            | 104           |
| 4497      | 1      | 515        | 269       | emb Z18852 SACF | S.aureus gene for clumping factor  | 99            | 213           |
| 4529      | 1      | 2          | 172       | emb X64172 SARP | S.aureus rplL, rpf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains | 100           | 151           |
| 4547      | 1      | 1          | 300       | emb X62992 SAFN | S.aureus fnbB gene for fibronectin binding protein B   | 100           | 157           |
| 4554      | 1      | 318        | 160       | emb Z18852 SACF | S.aureus gene for clumping factor  | 84            | 126           |
| 4555      | 1      | 9          | 237       | emb Z18852 SACF | S.aureus gene for clumping factor  | 84            | 213           |
| 4569      | 1      | 79         | 222       | emb Z18852 SACF | S.aureus gene for clumping factor  | 98            | 127           |
| 4608      | 1      | 22         | 216       | emb X58434 SAPD | S.aureus pphB, pphC and pphD genes for pyruvate decarboxylase, dihydroliipoamide acetyltransferase and dihydroliipoamide dehydrogenase                     | 92            | 168           |
| 4614      | 1      | 464        | 234       | emb Z18852 SACF | S.aureus gene for clumping factor  | 86            | 169           |
| 4623      | 1      | 105        | 302       | gb J04151       | S.aureus fibronectin-binding protein (fnbA) mRNA, complete cds   | 99            | 152           |
| 4632      | 1      | 18         | 206       | gb J03479       | S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds  | 98            | 183           |
| 4646      | 1      | 1          | 222       | emb Z18852 SACF | S.aureus gene for clumping factor  | 84            | 100           |
| 4687      | 1      | 2          | 166       | gb J04151       | S.aureus fibronectin-binding protein (fnbA) mRNA, complete cds   | 98            | 156           |

TABLE I

S. aureus - Coding regions containing known sequences

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | percent ident | HSP nt length | ORF nt length |
|-----------|--------|------------|-----------|-----------------|---|---------------|---------------|---------------|
| 4695      | 1      | 313        | 158       | g01114017       | Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds  | 75            | 155           | 156           |
| 4703      | 1      | 1          | 153       | emb/K58434/SAPD | S. aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase | 98            | 103           | 153           |

TABLE I

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 20        | 6      | 5089       | 4679      | gi 511839       | [ORF] [Staphylococcus bacteriophage phi 11]  | 100   | 100     | 411         |
| 149       | 3      | 2032       | 1577      | pir 849703 8497 | [int gene activator RlnA - bacteriophage phi 11]   | 100   | 100     | 456         |
| 149       | 5      | 2109       | 1912      | gi 166161       | [bacteriophage phi-11 int gene activator [Staphylococcus acteriophage phi 11]  | 100   | 100     | 198         |
| 349       | 2      | 558        | 409       | gi 166159       | [integrase (int) [Staphylococcus bacteriophage phi 11]   | 100   | 100     | 150         |
| 398       | 1      | 1372       | 707       | gi 166159       | [integrase (int) [Staphylococcus bacteriophage phi 11]   | 100   | 99      | 666         |
| 398       | 2      | 783        | 1001      | gi 455128       | [exciionase (xis) [Staphylococcus bacteriophage phi 11]  | 100   | 100     | 219         |
| 502       | 4      | 1916       | 1744      | gi 1204912      | [H. influenzae predicted coding region H10660 [Haemophilus influenzae]   | 100   | 71      | 171         |
| 849       | 1      | 2          | 262       | gi 1372002      | [polyprotein [Barn common mosaic virus]  | 100   | 46      | 261         |
| 1349      | 1      | 277        | 140       | gi 163359       | [protein synthesis initiation factor 2 (infB) [Bacillus subtilis] gl 49319   | 100   | 82      | 138         |
| 2880      | 1      | 21         | 308       | gi 862933       | [protein kinase C inhibitor-1 [Homo sapiens]   | 100   | 98      | 289         |
| 3085      | 1      | 428        | 216       | gi 1354211      | [PET112-like protein [Bacillus subtilis]   | 100   | 100     | 213         |
| 4168      | 2      | 571        | 398       | gi 1354211      | [PET112-like protein [Bacillus subtilis]   | 100   | 100     | 174         |
| 331       | 1      | 2          | 247       | gi 426473       | [nusG gene product [Staphylococcus carnosus]   | 98    | 95      | 246         |
| 307       | 2      | 1272       | 1463      | gi 460259       | [enolase [Bacillus subtilis]   | 97    | 90      | 192         |
| 331       | 2      | 395        | 850       | gi 581638       | [pil1 protein [Staphylococcus carnosus]  | 97    | 93      | 456         |
| 366       | 1      | 39         | 215       | gi 166161       | [bacteriophage phi-11 int gene activator [Staphylococcus acteriophage phi 11]  | 97    | 95      | 177         |
| 680       | 3      | 718        | 936       | gi 426473       | [nusG gene product [Staphylococcus carnosus]   | 97    | 97      | 219         |
| 3578      | 1      | 284        | 144       | gi 1339950      | [large subunit of NAMII-dependent glutamate synthase [Plectonura boryanum]   | 97    | 79      | 141         |
| 157       | 1      | 321        | 516       | gi 1022726      | [unknown [Staphylococcus haemolyticus]   | 96    | 88      | 198         |
| 205       | 33     | 18470      | 16147     | gi 1165302      | [S10 [Bacillus subtilis]   | 96    | 91      | 324         |
| 3919      | 1      | 48         | 401       | gi 1971784      | [Clp-like ATP-dependent protease binding subunit [Bos taurus]  | 96    | 81      | 354         |
| 4133      | 1      | 830        | 417       | gi 1022726      | [unknown [Staphylococcus haemolyticus]   | 96    | 84      | 414         |
| 4168      | 1      | 708        | 355       | gi 1354211      | [PET112-like protein [Bacillus subtilis]   | 96    | 95      | 354         |
| 4207      | 1      | 312        | 157       | gi 602031       | [similar to triethylamine DH [Mycoplasma capricolum] pir 849950 849950 probable trimethylamine dehydrogenase (EC 5.99.1) - Mycoplasma capricolum (SOC)] (fragment) | 96    | 86      | 156         |

TABLE 2

S. aureus - Pucative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start | Stop  | Match accession | Match gene name   | % sim | % ident | Length (nt) |
|-----------|--------|-------|-------|-----------------|---|-------|---------|-------------|
| 4227      | 2      | 152   | 331   | gi 871784       | Clp-like ATP-dependent protease binding subunit (Bos taurus)  | 96    | 81      | 180         |
| 4416      | 1      | 570   | 286   | gi 1022726      | unknown (Staphylococcus haemolyticus)   | 96    | 84      | 285         |
| 22        | 1      | 858   | 430   | gi 511070       | ureG (Staphylococcus xyloosus)  | 95    | 88      | 429         |
| 22        | 7      | 4362  | 4036  | gi 1581787      | urease gamma subunit (Staphylococcus xyloosus)  | 95    | 79      | 327         |
| 82        | 6      | 8794  | 9114  | gi JC0008 JG00  | ribosomal protein S7 - Bacillus stearothermophilus  | 95    | 83      | 321         |
| 154       | 9      | 9280  | 7838  | gi 1354211      | PEP112-like protein (Bacillus subtilis)   | 95    | 92      | 1443        |
| 186       | 3      | 2798  | 2055  | gi 1514636      | serine O-acetyltransferase (Staphylococcus xyloosus)  | 95    | 87      | 744         |
| 205       | 5      | 4406  | 4014  | gi 142462       | ribosomal protein S11 (Bacillus subtilis)   | 95    | 85      | 393         |
| 205       | 7      | 5017  | 4793  | gi 1142459      | initiation factor 1 (Bacillus subtilis)   | 95    | 84      | 225         |
| 205       | 31     | 11365 | 10991 | gi 1044974      | ribosomal protein L14 (Bacillus subtilis)   | 95    | 93      | 375         |
| 259       | 5      | 7288  | 6644  | sp P47995 VSEA  | HYPOTHETICAL PROTEIN IN SECA 5-REGION (ORF1) (FRAGMENT)   | 95    | 85      | 615         |
| 302       | 3      | 795   | 1097  | gi 40186        | homologous to E. coli ribosomal protein L27 (Bacillus subtilis) [113592 L27 ribosomal protein (Bacillus subtilis)] [r C1895 C21895 ribosomal protein L27 - Bacillus subtilis] [P03657 RL27_BACSU 50S RIBOSOMAL PROTEIN L27 (BL30) (BL4)]. [140175 L27 gene prod | 95    | 89      | 303         |
| 310       | 1      | 579   | 1523  | gi 1177684      | chorismate mutase (Staphylococcus xyloosus)   | 95    | 92      | 945         |
| 414       | 1      | 2     | 163   | gi C48396 C483  | ribosomal protein L34 - Bacillus stearothermophilus   | 95    | 90      | 162         |
| 4185      | 2      | 125   | 277   | gi 1276841      | glutamate synthase (GOGAT) (Porphyra purpurea)  | 95    | 86      | 153         |
| 22        | 2      | 1028  | 723   | gi 511069       | Ura1 (Staphylococcus xyloosus)  | 94    | 91      | 306         |
| 22        | 5      | 5046  | 3310  | gi 410516       | urease alpha subunit (Staphylococcus xyloosus)  | 94    | 85      | 1737        |
| 60        | 4      | 815   | 1372  | gi 666116       | glucose kinase (Staphylococcus xyloosus)  | 94    | 87      | 558         |
| 205       | 18     | 10012 | 9536  | gi 1044978      | ribosomal protein S8 (Bacillus subtilis)  | 94    | 78      | 477         |
| 326       | 4      | 3378  | 2342  | gi 557492       | [dihydroxynaphthoic acid (DHNA) synthetase (Bacillus subtilis)] [dihydroxynaphthoic acid (DHNA) synthetase (Bacillus subtilis)]   | 94    | 85      | 837         |
| 414       | 3      | 737   | 955   | gi 467386       | chlorophen and furan oxidation (Bacillus subtilis)  | 94    | 77      | 219         |
| 426       | 3      | 2260  | 1823  | gi 1263908      | putative (Staphylococcus epidermidis)   | 94    | 87      | 418         |
| 514       | 1      | 2     | 355   | gi 633650       | enzyme II (mannitol) (Staphylococcus carnosus)  | 94    | 84      | 354         |
| 1017      | 1      | 2     | 229   | gi 109435       | putative (Lactococcus lactis)   | 94    | 73      | 228         |
| 3098      | 1      | 330   | 184   | gi 413952       | lipo-28d gene product (Bacillus subtilis)   | 94    | 50      | 147         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 3232      | 1      | 630        | 316       | gi1022725       | unknown [Staphylococcus haemolyticus]  | 94    | 84      | 315         |
| 42        | 5      | 2089       | 2259      | pir848396 8483  | ribosomal protein L33 - Bacillus stearothermophilus  | 93    | 81      | 171         |
| 101       | 2      | 1745       | 1383      | gi155345        | arsenic efflux pump protein [Plasmid pSM267]   | 93    | 82      | 363         |
| 205       | 24     | 13227      | 11865     | ap14577 RL16    | 50S RIBOSOMAL PROTEIN L16  | 93    | 83      | 363         |
| 259       | 4      | 8291       | 5670      | gi1499335       | secA protein [Staphylococcus carnosus]   | 93    | 85      | 2619        |
| 275       | 1      | 2226       | 1114      | gi1631650       | enzyme II(mannitol) [Staphylococcus carnosus]  | 93    | 86      | 1113        |
| 444       | 6      | 6207       | 5773      | gi1022726       | unknown [Staphylococcus haemolyticus]  | 93    | 81      | 435         |
| 491       | 1      | 152        | 622       | gi146912        | ribosomal protein L13 [Staphylococcus carnosus]  | 93    | 88      | 471         |
| 607       | 6      | 1674       | 2033      | gi1022726       | unknown [Staphylococcus haemolyticus]  | 93    | 83      | 360         |
| 653       | 1      | 973        | 488       | gi1580890       | translation initiation factor IF3 (AA 1-172) [Bacillus stearothermophilus]   | 93    | 77      | 486         |
| 1864      | 1      | 3          | 194       | gi1306553       | ribosomal protein small subunit [Homo sapiens]   | 93    | 93      | 192         |
| 2997      | 1      | 28         | 300       | gi1143390       | carbamyl phosphate synthetase [Bacillus subtilis]  | 93    | 82      | 273         |
| 3232      | 2      | 907        | 596       | gi1022725       | unknown [Staphylococcus haemolyticus]  | 93    | 84      | 312         |
| 3761      | 2      | 794        | 621       | gi1022725       | unknown [Staphylococcus haemolyticus]  | 93    | 88      | 174         |
| 16        | 1      | 3          | 374       | gi1142781       | putative cytoplasmic protein; putative [Bacillus subtilis]<br>sp P37934 UVRB_BACSU EXCINUCLEASE ABC SUBUNIT B (DNA PROTEIN) FRAGMENT | 92    | 83      | 372         |
| 31        | 7      | 5915       | 6124      | gi1136430       | KIAA0185 protein [Homo sapiens]  | 92    | 46      | 210         |
| 56        | 19     | 26483      | 27391     | gi1467401       | unknown [Bacillus subtilis]  | 92    | 80      | 909         |
| 69        | 6      | 5882       | 6130      | gi1530200       | trophoblastin [Ovis aries]   | 92    | 53      | 249         |
| 145       | 3      | 2568       | 2038      | gi1022725       | unknown [Staphylococcus haemolyticus]  | 92    | 80      | 531         |
| 171       | 3      | 2760       | 2362      | gi1517475       | D-amino acid transaminase [Staphylococcus haemolyticus]  | 92    | 86      | 399         |
| 205       | 12     | 7495       | 6962      | gi149189        | secY gene product [Staphylococcus carnosus]  | 92    | 85      | 534         |
| 205       | 19     | 10812      | 10255     | gi1044976       | ribosomal protein L5 [Bacillus subtilis]   | 92    | 82      | 558         |
| 219       | 1      | 710        | 337       | gi1303812       | yqeV [Bacillus subtilis]   | 92    | 88      | 354         |
| 344       | 3      | 1575       | 1805      | gi11405474      | CspC protein [Bacillus cereus]   | 92    | 85      | 231         |
| 699       | 1      | 20         | 361       | gi1413999       | lpa-75d gene product [Bacillus subtilis]   | 92    | 81      | 302         |
| 1343      | 1      | 2          | 160       | pir145434 A454  | ribosomal protein L19 - Bacillus stearothermophilus  | 92    | 84      | 159         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 1958      | 1      | 524        | 264       | gi 407908       | [Eifac (Staphylococcus xyloosus)]  | 92    | 80      | 261         |
| 3578      | 2      | 718        | 386       | gi 1339950      | [large subunit of NAM-dependent glutamate synthase (Plectonema boryanum)]  | 92    | 78      | 333         |
| 3585      | 1      | 644        | 324       | gi 1339950      | [large subunit of NAM-dependent glutamate synthase (Plectonema boryanum)]  | 92    | 81      | 321         |
| 3640      | 3      | 4          | 402       | gi 1022726      | [unknown (Staphylococcus haemolyticus)]  | 92    | 81      | 399         |
| 4362      | 1      | 14         | 178       | gi 450688       | [hcdw gene of Ecopirri gene product (Escherichia coli) pir[SJ8437/SJ8437 hcdw protein - Escherichia coli pir[SJ8437/SJ8437 hypothetical protein A - Escherichia coli (SUB 40-520)] | 92    | 76      | 165         |
| 4446      | 1      | 338        | 182       | gi 1022725      | [unknown (Staphylococcus haemolyticus)]  | 92    | 82      | 177         |
| 4549      | 1      | 462        | 232       | gi 1022726      | [unknown (Staphylococcus haemolyticus)]  | 92    | 80      | 231         |
| 4626      | 1      | 3          | 224       | gi 1022725      | [unknown (Staphylococcus haemolyticus)]  | 92    | 84      | 222         |
| 2         | 4      | 3980       | 4531      | gi 535349       | [Codw (Bacillus subtilis)]   | 91    | 74      | 552         |
| 28        | 1      | 2          | 1126      | gi 1001376      | [hypothetical protein (Synechocystis sp.)]   | 91    | 78      | 1125        |
| 60        | 5      | 1354       | 1701      | gi 1226043      | [orf2 downstream of glucose kinase (Staphylococcus xyloosus)]  | 91    | 80      | 348         |
| 101       | 1      | 1989       | 1036      | gi 150728       | [arsenic efflux pump protein (Plasmid p1358)]  | 91    | 80      | 954         |
| 187       | 2      | 412        | 1194      | gi 142559       | [ATP synthase alpha subunit (Bacillus megaterium)]   | 91    | 79      | 783         |
| 205       | 22     | 11579      | 11298     | gi 40149        | [S17 protein (AA 1-87) (Bacillus subtilis)]  | 91    | 83      | 282         |
| 206       | 7      | 8184       | 10262     | gi 1072418      | [glcA gene product (Staphylococcus carnosus)]  | 91    | 83      | 2079        |
| 306       | 2      | 3885       | 2326      | gi 143012       | [GMP synthetase (Bacillus subtilis)]   | 91    | 78      | 1560        |
| 306       | 3      | 5319       | 3826      | gi 467399       | [IMP dehydrogenase (Bacillus subtilis)]  | 91    | 79      | 1494        |
| 310       | 3      | 2194       | 3207      | gi 1177685      | [ccpA gene product (Staphylococcus xyloosus)]  | 91    | 81      | 1014        |
| 343       | 4      | 2974       | 3150      | gi 949974       | [sucrose repressor (Staphylococcus xyloosus)]  | 91    | 82      | 177         |
| 480       | 3      | 1606       | 3042      | gi 433991       | [ATP synthase subunit beta (Bacillus subtilis)]  | 91    | 85      | 1437        |
| 536       | 3      | 2026       | 1280      | gi 143166       | [adenylosuccinate lyase (Pur-8) (Bacillus subtilis) pir[C29326/W25SDS adenylosuccinate lyase (EC 4.3.2.2) - Bacillus subtilis]   | 91    | 79      | 747         |
| 552       | 1      | 1064       | 615       | gi 297874       | [fructose-bisphosphate aldolase (Staphylococcus carnosus) pir[A49943/A49943 fructose-bisphosphate aldolase (EC 4.1.2.13) - Staphylococcus carnosus (strain TM300)]                 | 91    | 79      | 450         |
| 637       | 1      | 1          | 1536      | gi 143597       | [GTP synthetase (Bacillus subtilis)]   | 91    | 79      | 1536        |
| 859       | 1      | 21         | 359       | gi 385178       | [unknown (Bacillus subtilis)]  | 91    | 66      | 339         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 327       | 1      | 339        | 530       | gi 496558       | orfX (Bacillus subtilis)   | 91    | 71      | 192         |
| 2515      | 1      | 466        | 275       | gi 51070        | orfG (Staphylococcus xylosum)  | 91    | 85      | 192         |
| 2594      | 1      | 2          | 202       | gi 146824       | beta-cystathionase (Bacterichia coli)  | 91    | 75      | 201         |
| 3764      | 1      | 847        | 425       | gi 1022725      | unknown (Staphylococcus haemolyticus)  | 91    | 78      | 423         |
| 4011      | 1      | 127        | 495       | gi 1022726      | unknown (Staphylococcus haemolyticus)  | 91    | 75      | 369         |
| 4227      | 1      | 1          | 177       | gi 296464       | ATPase (Lactococcus lactis)  | 91    | 66      | 177         |
| 42        | 3      | 815        | 1033      | gi 520401       | catalase (Haemophilus influenzae)  | 90    | 86      | 219         |
| 51        | 8      | 3717       | 4607      | gi 580899       | OppF gene product (Bacillus subtilis)  | 90    | 74      | 891         |
| 129       | 3      | 5317       | 4001      | gi 1146206      | glutamate dehydrogenase (Bacillus subtilis)  | 90    | 76      | 1317        |
| 164       | 17     | 16628      | 16933     | ep P05766 RS15  | 30S RIBOSOMAL PROTEIN S15 (BS18)   | 90    | 74      | 306         |
| 171       | 5      | 2983       | 2819      | gi 517475       | D-amino acid transaminase (Staphylococcus haemolyticus)  | 90    | 76      | 165         |
| 205       | 4      | 4497       | 3550      | gi 142463       | RNA polymerase alpha-core-subunit (Bacillus subtilis)  | 90    | 76      | 948         |
| 205       | 6      | 4748       | 4410      | gi 1044989      | ribosomal protein S13 (Bacillus subtilis)  | 90    | 73      | 319         |
| 205       | 10     | 7165       | 6104      | gi 49189        | secY gene product (Staphylococcus carnosus)  | 90    | 81      | 762         |
| 205       | 12     | 6645       | 6472      | gi 49189        | secY gene product (Staphylococcus carnosus)  | 90    | 78      | 174         |
| 205       | 27     | 13692      | 13345     | gi 786157       | Ribosomal Protein S19 (Bacillus subtilis)  | 90    | 79      | 348         |
| 205       | 31     | 15858      | 15496     | gi 1165303      | l3 (Bacillus subtilis)   | 90    | 79      | 363         |
| 260       | 5      | 7023       | 5773      | gi 1161380      | IcaA (Staphylococcus epidermidis)  | 90    | 78      | 1251        |
| 299       | 6      | 3376       | 3947      | gi 467440       | phosphoribosylpyrophosphate synthetase (Bacillus subtilis) gi 40218 PRPP synthetase (AA 1-317) (Bacillus subtilis)                 | 90    | 78      | 570         |
| 320       | 2      | 1025       | 1217      | gi 312443       | carbamoyl-phosphate synthase (glutamine-hydrolyzing) (Bacillus aldolyticus)  | 90    | 75      | 693         |
| 330       | 4      | 1581       | 1769      | gi 986963       | beta-tubulin (Sporidobolus parvovirus)   | 90    | 80      | 189         |
| 369       | 1      | 954        | 523       | gi 534763 5347  | L-serine dehydratase beta chain - Clostridium sp.  | 90    | 77      | 432         |
| 557       | 1      | 3          | 188       | gi 151589       | M. jannaschii predicted coding region M1624 (Methanococcus jannaschii)   | 90    | 54      | 186         |
| 663       | 2      | 667        | 1200      | gi 143786       | tryptophanyl-tRNA synthetase (EC 6.1.1.2) (Bacillus subtilis) pi J57048 TWS cryptophan-tRNA ligase (EC 6.1.1.2) - Bacillus ubtilla | 90    | 73      | 534         |
| 717       | 1      | 1          | 261       | gi 143065       | hubat (Bacillus stearothermophilus)  | 90    | 79      | 261         |
| 745       | 4      | 1059       | 865       | gi 1205433      | H. influenzae predicted coding region H1190 (Haemophilus influenzae)   | 90    | 81      | 195         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 1007      | 1      | 386        | 565       | gi 1143366      | adenylosuccinate lyase (PUB-9) [Bacillus subtilis] p1r[C29326]W23SDS                               | 90    | 77      | 180         |
| 1054      | 1      | 579        | 731       | gi 1103322      | adenylosuccinate lyase (EC 4.3.2.2) - Bacillus subtilis  |       |         |             |
| 1156      | 1      | 117        | 707       | gi 1147776      | ORF_729 [Escherichia coli]   | 90    | 50      | 249         |
| 1180      | 1      | 408        | 205       | gi 1177831      | ClpP [Bacillus subtilis]   | 90    | 80      | 591         |
| 1253      | 1      | 1          | 462       | gi 40046        | unknown [Bacillus subtilis]  | 90    | 74      | 204         |
|           |        |            |           |                 | phosphoglucose isomerase A (AA 1-449) [Bacillus stearothermophilus]                                |       |         |             |
|           |        |            |           |                 | tr[S15936]NUB6SA glucose-6-phosphate isomerase (EC 5.3.1.9) A - cillus stearothermophilus          | 90    | 75      | 462         |
| 2951      | 1      | 3          | 269       | gi 1144816      | formyltetrahydrofolate synthetase (FTHFS) (cog start codon) (EC 3.4.3) [Mycobacteria thermocellus] | 90    | 76      | 267         |
| 3160      | 1      | 327        | 166       | gi 1070014      | protein-dependent [Bacillus subtilis]  | 90    | 52      | 162         |
| 4594      | 1      | 3          | 233       | gi 1871784      | Clp-like ATP-dependent protease binding subunit [Bos taurus]                                       | 90    | 76      | 231         |
| 87        | 1      | 1028       | 1750      | gi 1467327      | unknown [Bacillus subtilis]  | 89    | 75      | 723         |
| 112       | 1      | 2          | 505       | gi 153741       | ATP-binding protein [Streptococcus mutans]   | 89    | 77      | 504         |
| 118       | 1      | 120        | 398       | gi 1303804      | YqeQ [Bacillus subtilis]   | 89    | 75      | 279         |
| 126       | 4      | 3545       | 3757      | gi 1460257      | triose phosphate isomerase [Bacillus subtilis]   | 89    | 84      | 211         |
| 164       | 12     | 11667      | 12755     | gi 39954        | IP2 (aa 1-741) [Bacillus stearothermophilus]   | 89    | 80      | 1089        |
| 205       | 13     | 7875       | 7405      | gi 216138       | ORF for L15 ribosomal protein [Bacillus subtilis]  | 89    | 76      | 471         |
| 205       | 32     | 16132      | 13623     | gi 1165303      | L3 [Bacillus subtilis]   | 89    | 80      | 330         |
| 270       | 3      | 2407       | 2207      | gi C61902[C419] | arsenate reductase (EC 1.1.1.2) - Staphylococcus xyloosus plasmid pSX267                           | 89    | 81      | 201         |
| 395       | 2      | 157        | 872       | gi 520574       | glutamate racemase [Staphylococcus haemolyticus]   | 89    | 80      | 516         |
| 498       | 1      | 3          | 839       | gi 396259       | protease [Staphylococcus epidermidis]  | 89    | 77      | 837         |
| 510       | 1      | 1          | 464       | gi 40046        | phosphoglucose isomerase A (AA 1-449) [Bacillus stearothermophilus]                                | 89    | 74      | 444         |
|           |        |            |           |                 | tr[S15936]NUB6SA glucose-6-phosphate isomerase (EC 5.3.1.9) A - cillus stearothermophilus          |       |         |             |
| 615       | 1      | 2124       | 1210      | gi 1303812      | YqeV [Bacillus subtilis]   | 89    | 74      | 915         |
| 841       | 1      | 18         | 361       | gi 1165303      | L3 [Bacillus subtilis]   | 89    | 80      | 324         |
| 1111      | 1      | 352        | 813       | gi 47146        | thermonuclease [Staphylococcus intermedius]  | 89    | 70      | 462         |
| 1875      | 1      | 2          | 256       | gi 1205108      | ATP-dependent protease binding subunit [Haemophilus influenzae]                                    | 89    | 82      | 255         |
| 2963      | 1      | 11         | 367       | gi 467458       | cell division protein [Bacillus subtilis]  | 89    | 83      | 357         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match / 1 | gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|-----------|--|-------|---------|-------------|
| 3020      | 1      | 90         | 362       | gi1239988       |           | hypothetical protein [Bacillus subtilis]   | 89    | 66      | 273         |
| 3565      | 1      | 2          | 400       | gi1256635       |           | dihydroxy-acid dehydratase [Bacillus subtilis]   | 89    | 75      | 399         |
| 3586      | 1      | 105        | 314       | gi380832        |           | ATP synthase subunit gamma [Bacillus subtilis]   | 89    | 82      | 210         |
| 3629      | 1      | 794        | 399       | gi1009366       |           | Respiratory nitrate reductase [Bacillus subtilis]  | 89    | 78      | 396         |
| 3688      | 1      | 2          | 400       | gi1146206       |           | glutamate dehydrogenase [Bacillus subtilis]  | 89    | 75      | 399         |
| 3699      | 1      | 794        | 399       | gi1133950       |           | large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]   | 89    | 75      | 396         |
| 4016      | 1      | 428        | 216       | gi1009366       |           | Respiratory nitrate reductase [Bacillus subtilis]  | 89    | 71      | 213         |
| 4177      | 1      | 471        | 301       | gi1149426       |           | putative [Lactococcus lactis]  | 89    | 76      | 171         |
| 4436      | 1      | 601        | 302       | gi1022725       |           | unknown [Staphylococcus haemolyticus]  | 89    | 80      | 300         |
| 4635      | 1      | 320        | 162       | gi1022725       |           | unknown [Staphylococcus haemolyticus]  | 89    | 73      | 159         |
| 2         | 2      | 1330       | 2676      | gi1520754       |           | putative [Bacillus subtilis]   | 88    | 76      | 1347        |
| 42        | 2      | 468        | 848       | sp P42321 CATA  |           | CATALASE [EC 1.11.1.6]   | 88    | 76      | 381         |
| 53        | 5      | 6189       | 4722      | gi1474177       |           | alpha-D-1,4-glucosidase [Staphylococcus xylosum]   | 88    | 80      | 1668        |
| 56        | 16     | 18018      | 18617     | gi1467411       |           | recombination protein [Bacillus subtilis]  | 88    | 77      | 600         |
| 60        | 3      | 376        | 843       | gi1466116       |           | glucose kinase [Staphylococcus xylosum]  | 88    | 77      | 468         |
| 70        | 2      | 1583       | 1245      | gi146095        |           | replication initiator protein [Listeria monocytogenes]   | 88    | 74      | 339         |
| 82        | 8      | 11514      | 12719     | gi1460663       |           | translation elongation factor tu - Bacillus subtilis   | 88    | 79      | 1206        |
| 103       | 7      | 4179       | 4391      | gi1167181       |           | serine/threonine kinase receptor [Brassica napus]  | 88    | 77      | 213         |
| 114       | 8      | 7732       | 8232      | gi1022726       |           | unknown [Staphylococcus haemolyticus]  | 88    | 72      | 501         |
| 118       | 2      | 308        | 2011      | gi11303804      |           | XqeQ [Bacillus subtilis]   | 88    | 77      | 1704        |
| 141       | 3      | 657        | 1136      | gi11403446      |           | transketolase [Bacillus subtilis]  | 88    | 72      | 480         |
| 148       | 7      | 5871       | 6116      | gi1118002       |           | dihydroterate synthase [Staphylococcus haemolyticus]   | 88    | 78      | 246         |
| 165       | 3      | 1428       | 2231      | gi140053        |           | phenylalanyl-tRNA synthetase alpha subunit [Bacillus subtilis]<br>tr S11730 YFSA phenylalanine-tRNA ligase [EC 6.1.1.20] alpha subunit - Bacillus subtilis | 88    | 80      | 804         |
| 205       | 28     | 15027      | 14185     | gi1165306       |           | L2 [Bacillus subtilis]   | 88    | 82      | 843         |
| 225       | 1      | 1569       | 898       | gi11203840      |           | Xqfs [Bacillus subtilis]   | 88    | 78      | 672         |
| 235       | 1      | 2          | 1975      | gi1452109       |           | valyl-tRNA synthetase [Bacillus subtilis]  | 88    | 76      | 1974        |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 339       | 3      | 2040       | 1566      | gi 1118002      | [dihydropterate synthase (Staphylococcus haemolyticus)]   | 88    | 73      | 495         |
| 443       | 4      | 4325       | 2928      | gi 558559       | [pyrimidine nucleoside phosphorylase (Bacillus subtilis)]   | 88    | 73      | 1398        |
| 532       | 1      | 3          | 419       | gi 143797       | [valyl-tRNA synthetase (Bacillus stearothermophilus) sp P19311SV_BACST VALYL-TRNA SYNTHETASE (EC 6.1.1.9) VALINE--TRNA LIGASE) (VALANS)]  | 88    | 78      | 417         |
| 534       | 3      | 2504       | 2968      | gi 151049       | [mannitol-specific enzyme-III (Staphylococcus carnosus) p J00088 J00088 phosphotransferase system enzyme III (EC 7.1.4.9) mannitol-specific factor III - Staphylococcus carnosus (p J1876 PTNA_STACA PIS SYSTEM, MANNITOL-SPECIFIC IIA COMPONENT E1A-MTL) I | 88    | 82      | 465         |
| 705       | 2      | 584        | 399       | gi 710018       | [nitrite reductase (nirB) (Bacillus subtilis)]  | 88    | 70      | 186         |
| 1000      | 2      | 1824       | 1309      | gi 1022726      | [unknown (Staphylococcus haemolyticus)]   | 88    | 78      | 516         |
| 1299      | 1      | 587        | 324       | gi 401786       | [phosphomannomutase (Mycoplasma pirum)]   | 88    | 55      | 264         |
| 1341      | 2      | 170        | 400       | gi 39963        | [ribosomal protein L20 (M1-119) (Bacillus stearothermophilus) p J50314 J50310 ribosomal protein L20 - Bacillus stearothermophilus   | 88    | 82      | 231         |
| 1386      | 1      | 41         | 214       | p J847154 B471  | [signal recognition particle 54K chain homolog Ffh - Bacillus subtilis]   | 88    | 71      | 174         |
| 1386      | 2      | 183        | 533       | p J847154 B471  | [signal recognition particle 54K chain homolog Ffh - Bacillus subtilis]   | 88    | 73      | 351         |
| 2949      | 3      | 704        | 399       | gi 535150       | [CodX (Bacillus subtilis)]  | 88    | 73      | 306         |
| 2984      | 1      | 5          | 169       | gi 218277       | [O-acetylserine(thiol) lyase (Spinacia oleracea)]   | 88    | 70      | 165         |
| 3035      | 1      | 1          | 138       | gi 493083       | [dihydroxyacetone kinase (Citrobacter freundii)]  | 88    | 67      | 138         |
| 3089      | 1      | 3          | 152       | gi 606055       | [ORF F746 (Escherichia coli)]   | 88    | 88      | 150         |
| 3917      | 1      | 817        | 410       | gi 143378       | [pyruvate decarboxylase (E-1) beta subunit (Bacillus subtilis) gi 1377836 pyruvate decarboxylase E-1 beta subunit (Bacillus subtilis)]  | 88    | 77      | 408         |
| 4199      | 1      | 680        | 342       | gi 1405454      | [aconitase (Bacillus subtilis)]   | 88    | 82      | 339         |
| 4201      | 1      | 734        | 369       | gi 515938       | [glutamate synthase (ferredoxin) (Synechocystis sp.) p J546957 J546957 glutamate synthase (ferredoxin) (EC 1.4.7.1) - Synechocystis sp.   | 88    | 84      | 366         |
| 4274      | 1      | 1          | 336       | gi 515938       | [glutamate synthase (ferredoxin) (Synechocystis sp.) p J546957 J546957 glutamate synthase (ferredoxin) (EC 1.4.7.1) - Synechocystis sp.   | 88    | 84      | 336         |
| 4308      | 1      | 794        | 399       | gi 1146206      | [glutamate dehydrogenase (Bacillus subtilis)]   | 88    | 71      | 396         |
| 2         | 5      | 4570       | 6000      | gi 533350       | [CodX (Bacillus subtilis)]  | 87    | 70      | 1431        |
| 52        | 8      | 6781       | 6482      | gi 1084791      | [function unknown (Bacillus subtilis)]  | 87    | 66      | 300         |

TABLE 2

S. aureus - putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 73        | 3      | 1584       | 2480      | gi1142952       | glycerol kinase (gpk) (EC 2.7.1.30) [Bacillus subtilis] pir[B45868]B45868<br>glycerol kinase (EC 2.7.1.30) - Bacillus subtilis sp[B18157]B18157<br>GLYCEROL KINASE (EC 2.7.1.30) (ATP:GLYCEROL - PHOSPHOTRANSFERASE) (GLYCEROKINASE) (OK) | 87    | 72      | 897         |
| 98        | 12     | 8813       | 9100      | gi11467433      | unknown [Bacillus subtilis]   | 87    | 62      | 288         |
| 124       | 4      | 4265       | 2988      | gi1556886       | serine hydroxymethyltransferase [Bacillus subtilis] pir[S49363]S49363<br>serine hydroxymethyltransferase - Bacillus ubctilis  | 87    | 77      | 1278        |
| 124       | 6      | 4457       | 4012      | gi1556883       | unknown [Bacillus subtilis]   | 87    | 66      | 426         |
| 148       | 5      | 3741       | 4559      | gi1467460       | unknown [Bacillus subtilis]   | 87    | 70      | 819         |
| 164       | 13     | 12710      | 13810     | gi1399554       | [PF2 (aa 1-741) [Bacillus stearothermophilus]   | 87    | 72      | 1101        |
| 177       | 2      | 1104       | 2126      | gi1467385       | unknown [Bacillus subtilis]   | 87    | 78      | 1023        |
| 199       | 1      | 1982       | 1158      | gi1141527       | [iron-sulfur protein [Bacillus subtilis]  | 87    | 77      | 825         |
| 199       | 2      | 4717       | 2933      | pir[A27763]A277 | [succinate dehydrogenase (EC 1.3.99.1) flavoprotein - Bacillus subtilis]  | 87    | 80      | 1785        |
| 205       | 23     | 11782      | 11543     | gi11044972      | [ribosomal protein L29 [Bacillus subtilis]  | 87    | 78      | 240         |
| 205       | 25     | 11325      | 12607     | gi1165309       | [S3 [Bacillus subtilis]   | 87    | 75      | 669         |
| 222       | 1      | 2033       | 1107      | gi1117249       | [rec23 gene product [Bacillus subtilis]   | 87    | 70      | 927         |
| 236       | 3      | 1635       | 1333      | gi1146198       | [ferredoxin [Bacillus subtilis]   | 87    | 80      | 303         |
| 246       | 5      | 2585       | 2292      | gi1467373       | [ribosomal protein S18 [Bacillus subtilis]  | 87    | 77      | 294         |
| 260       | 2      | 4189       | 3422      | gi11161382      | [fcsC [Staphylococcus epidermidis]  | 87    | 72      | 768         |
| 320       | 3      | 1696       | 2391      | gi1312443       | [carbamoyl-phosphate synthase (glutamine-hydrolyzing) [Bacillus aldolycicus]  | 87    | 80      | 696         |
| 380       | 4      | 1165       | 1383      | gi1142570       | [ATP synthase c subunit [Bacillus firmus]   | 87    | 80      | 219         |
| 414       | 4      | 900        | 1073      | gi1467386       | [thiophen and furan oxidation [Bacillus subtilis]   | 87    | 77      | 174         |
| 425       | 2      | 1003       | 794       | gi11046166      | [pilin repressor [Mycoplasma genitalium]  | 87    | 69      | 210         |
| 448       | 1      | 1255       | 722       | gi1405134       | [acetate kinase [Bacillus subtilis]   | 87    | 75      | 534         |
| 480       | 1      | 1          | 711       | gi1142559       | [ATP synthase alpha subunit [Bacillus megaterium]   | 87    | 79      | 711         |
| 481       | 1      | 2          | 352       | sw1006797[RL]_B | [50S RIBOSOMAL PROTEIN L1 (RL1)]  | 87    | 72      | 351         |
| 677       | 2      | 359        | 955       | gi1460911       | [fructose-bisphosphate aldolase [Bacillus subtilis]   | 87    | 78      | 597         |
| 677       | 3      | 934        | 1284      | gi1460911       | [fructose-bisphosphate aldolase [Bacillus subtilis]   | 87    | 78      | 351         |
| 876       | 1      | 3          | 452       | gi1146247       | [asparaginyl-tRNA synthetase [Bacillus subtilis]  | 87    | 79      | 450         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match      | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|------------|---|-------|---------|-------------|
| 1376      | 1      | 426        | 214       | gi1065555       | gi1065555  | [F4H6.4 gene product (Caenorhabditis elegans)]  | 87    | 75      | 213         |
| 2206      | 1      | 3          | 374       | gi213098        | gi213098   | [excisionase (Bacteriophage 154a)]  | 87    | 72      | 372         |
| 2918      | 1      | 3          | 290       | gi508979        | gi508979   | [GTP-binding protein (Bacillus subtilis)]   | 87    | 69      | 288         |
| 3081      | 2      | 126        | 308       | gi1467199       | gi1467199  | [IMP dehydrogenase (Bacillus subtilis)]   | 87    | 72      | 183         |
| 3535      | 1      | 3          | 401       | gi1405454       | gi1405454  | [aconitase (Bacillus subtilis)]   | 87    | 80      | 399         |
| 4218      | 1      | 547        | 275       | gi1603769       | gi1603769  | [Hutu protein, urocanase (Bacillus subtilis)]   | 87    | 73      | 273         |
| 4         | 8      | 10427      | 8736      | gi1603769       | gi1603769  | [Hutu protein, urocanase (Bacillus subtilis)]   | 86    | 72      | 1692        |
| 22        | 6      | 4190       | 3738      | gi1410515       | gi1410515  | [urease beta subunit (Staphylococcus xylosum)]  | 86    | 73      | 453         |
| 54        | 2      | 2480       | 1572      | gi1289287       | gi1289287  | [UDP-glucose pyrophosphorylase (Bacillus subtilis)]   | 86    | 70      | 909         |
| 124       | 3      | 2336       | 1713      | gi1556887       | gi1556887  | [uracil phosphoribosyltransferase (Bacillus subtilis)]  | 86    | 74      | 624         |
| 148       | 3      | 1349       | 3448      | gi1467458       | gi1467458  | [cell division protein (Bacillus subtilis)]   | 86    | 75      | 2100        |
| 148       | 4      | 3638       | 3859      | gi1467460       | gi1467460  | [unknown (Bacillus subtilis)]   | 86    | 73      | 222         |
| 352       | 3      | 1340       | 2086      | gi1377835       | gi1377835  | [pyruvate decarboxylase E-1 alpha subunit (Bacillus subtilis)]                                    | 86    | 75      | 747         |
| 164       | 18     | 17347      | 19467     | gi1184680       | gi1184680  | [polynucleotide phosphorylase (Bacillus subtilis)]  | 86    | 72      | 2121        |
| 180       | 2      | 554        | 1159      | gi143467        | gi143467   | [ribosomal protein S4 (Bacillus subtilis)]  | 86    | 80      | 606         |
| 205       | 3      | 2966       | 2592      | gi142464        | gi142464   | [ribosomal protein L17 (Bacillus subtilis)]   | 86    | 77      | 375         |
| 205       | 26     | 13364      | 12990     | gi140107        | gi140107   | [ribosomal protein L22 (Bacillus stearothermophilus)]   | 86    | 75      | 375         |
| 246       | 7      | 3463       | 3140      | gi1467375       | gi1467375  | [ribosomal protein S6 (Bacillus subtilis)]  | 86    | 70      | 324         |
| 299       | 3      | 1196       | 1340      | gi139656        | gi139656   | [spoVG gene product (Bacillus megaterium)]  | 86    | 70      | 345         |
| 299       | 7      | 3884       | 4345      | gi1467440       | gi1467440  | [phosphoribosylpyrophosphate synthetase (Bacillus subtilis)]                                      | 86    | 78      | 462         |
| 304       | 5      | 2170       | 2523      | gi1666983       | gi1666983  | [putative ATP binding subunit (Bacillus subtilis)]  | 86    | 65      | 354         |
| 310       | 2      | 1487       | 1678      | gi1177684       | gi1177684  | [chorismate mutase (Staphylococcus xylosum)]  | 86    | 71      | 192         |
| 337       | 5      | 2086       | 3405      | gi1467434       | gi1467434  | [isocitrate dehydrogenase (Bacillus subtilis)]  | 86    | 78      | 1320        |
| 339       | 2      | 1489       | 1109      | gi11118003      | gi11118003 | [dihydropyrimidin aldolase (Staphylococcus haemolyticus)]   | 86    | 77      | 381         |
| 358       | 2      | 2124       | 3440      | gi1146219       | gi1146219  | [38.2% of identity to the Escherichia coli GTP-binding protein Era; putative (Bacillus subtilis)] | 86    | 73      | 1317        |

TABLE 2

5 aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession  | Match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|--|---|-------|---------|-------------|
| 404       | 2      | 1015       | 2058      | [gi1103817]  | [yqfA (Bacillus subtilis)]  | 86    | 78      | 1044        |
| 581       | 2      | 861        | 452       | [gi140056]   | [phoP gene product (Bacillus subtilis)]                                 | 86    | 71      | 210         |
| 642       | 2      | 338        | 1075      | [gi11176399]   | [spiF (Staphylococcus epidermidis)]                                     | 86    | 72      | 738         |
| 770       | 1      | 422        | 347       | [gi143328]   | [phoP protein (put.) putative (Bacillus subtilis)]                      | 86    | 69      | 276         |
| 865       | 1      | 1777       | 890       | [gi11146247]   | [asparaginyl-tRNA synthetase (Bacillus subtilis)]                       | 86    | 74      | 888         |
| 868       | 2      | 963        | 1133      | [gi1002911]  | [transmembrane protein (Saccharomyces cerevisiae)]                      | 86    | 69      | 171         |
| 904       | 1      | 1          | 162       | [nt1303912]  | [yghW (Bacillus subtilis)]  | 86    | 72      | 162         |
| 989       | 1      | 35         | 433       | [gi1303993]  | [yqkL (Bacillus subtilis)]  | 86    | 76      | 399         |
| 1212      | 1      | 296        | 150       | [gi1414014]  | [ipa-90d gene product (Bacillus subtilis)]                              | 86    | 70      | 147         |
| 1323      | 1      | 2          | 148       | [nt140041]   | [pyruvate dehydrogenase (liponamide) (Bacillus stearothermophilus)]     | 86    | 75      | 147         |
|           |        |            |           | [ir1810798] [BMSPP pyruvate dehydrogenase (liponamide) (EC 1.2.4.1) pha chain - Bacillus stearothermophilus] |   |       |         |             |
| 3085      | 2      | 540        | 310       | [gi1354211]  | [PMT12-like protein (Bacillus subtilis)]                                | 86    | 86      | 231         |
| 1847      | 1      | 1          | 228       | [nt1296464]  | [ATPase (Lactococcus lactis)]   | 86    | 63      | 228         |
| 4487      | 1      | 476        | 240       | [gi1022726]  | [unknown (Staphylococcus haemolyticus)]                                 | 86    | 73      | 237         |
| 4503      | 1      | 372        | 187       | [gi1022725]  | [unknown (Staphylococcus haemolyticus)]                                 | 86    | 79      | 186         |
| 25        | 5      | 4287       | 5039      | [gi1502421]  | [3-ketoacyl-acyl carrier protein reductase (Bacillus subtilis)]         | 85    | 64      | 753         |
| 56        | 121    | 30627      | 32935     | [gi1408507]  | [pyrimidine nucleoside transport protein (Bacillus subtilis)]           | 85    | 69      | 1213        |
| 68        | 2      | 332        | 1192      | [gi1467376]  | [unknown (Bacillus subtilis)]   | 85    | 74      | 861         |
| 73        | 2      | 880        | 1707      | [gi142992]   | [glycerol kinase (GlpK) (EC 2.7.1.30) (Bacillus subtilis)]              | 85    | 72      | 828         |
|           |        |            |           | [glycerol kinase (EC 2.7.1.30) - Bacillus subtilis sp18157] [GLPK_BACSU]                                     |   |       |         |             |
|           |        |            |           | [GLYCEROL KINASE (EC 2.7.1.30) (ATP:GLYCEROL -PHOSPHOTRANSFERASE) (GLYCEROKINASE) (GK)]                      |   |       |         |             |
| 106       | 4      | 1505       | 3490      | [gi143766]   | [GhrSV (EC 6.1.1.3) (Bacillus subtilis)]                                | 85    | 74      | 1986        |
| 124       | 2      | 1153       | 2202      | [gi131924]   | [glycerol dehydro-3-phosphate dehydrogenase (Clostridium pasteurianum)] | 85    | 75      | 1050        |
|           |        |            |           | [pir151234] [S34256 glycerol dehydro-3-phosphate dehydrogenase (EC 2.1.1.12) - Clostridium pasteurianum]     |   |       |         |             |
| 129       | 4      | 6466       | 5252      | [gi1064807]  | [ORTHONINE AMINOTRANSFERASE (Bacillus subtilis)]                        | 85    | 73      | 1215        |
| 138       | 6      | 3475       | 5673      | [gi1072419]  | [glcB gene product (Staphylococcus carnosus)]                           | 85    | 74      | 2199        |
| 189       | 1      | 2          | 169       | [gi1467385]  | [unknown (Bacillus subtilis)]   | 85    | 65      | 168         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession  | Match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|------------------|--|-------|---------|-------------|
| 205       | 15     | 8624       | 8106      | [gi 104498]      | [ribosomal protein S5 (Bacillus subtilis)]   | 85    | 75      | 519         |
| 205       | 20     | 10928      | 10596     | [gi A02819 SBS]  | [ribosomal protein L24 - Bacillus acetotheophilus]   | 85    | 72      | 333         |
| 220       | 6      | 6490       | 6101      | [gi 48980]       | [secA gene product (Bacillus subtilis)]  | 85    | 66      | 390         |
| 231       | 4      | 4877       | 3159      | [gi 1002520]     | [MutS (Bacillus subtilis)]   | 85    | 70      | 1719        |
| 243       | 9      | 8013       | 8783      | [gi 414011]      | [lpa-87r gene product (Bacillus subtilis)]   | 85    | 72      | 771         |
| 249       | 2      | 5894       | 3186      | [gi 1405454]     | [aconitase (Bacillus subtilis)]  | 85    | 73      | 2709        |
| 302       | 1      | 140        | 475       | [gi 40173]       | [homolog of E. coli ribosomal protein L21 (Bacillus subtilis)]   | 85    | 72      | 336         |
|           |        |            |           |                  | [ir S18439 S18439 Ribosomal protein L21 - Bacillus subtilis p P26908 RL21_BACSU 50S RIBOSOMAL PROTEIN L21 (BL20)]                              |       |         |             |
| 333       | 1      | 5445       | 2968      | [gi 442360]      | [ClpC adenosine triphosphatase (Bacillus subtilis)]  | 85    | 69      | 2478        |
| 364       | 6      | 6082       | 8196      | [gi 871784]      | [Clp-like ATP-dependent protease binding subunit (Bos taurus)]   | 85    | 68      | 2115        |
| 448       | 2      | 1992       | 1339      | [gi 405134]      | [acetate kinase (Bacillus subtilis)]   | 85    | 68      | 654         |
| 747       | 1      | 1251       | 853       | [gi 1173157]     | [orf-X; hypothetical protein; Method: conceptual translation supplied by author (Bacillus subtilis)]   | 85    | 73      | 399         |
| 886       | 2      | 159        | 467       | [gi 541768]      | [hemin permease (Yersinia enterocolitica)]   | 85    | 55      | 309         |
| 1049      | 1      | 1208       | 606       | [gi 847154 8471] | [signal recognition particle 51K chain homolog F/h - Bacillus subtilis]  | 85    | 71      | 603         |
| 1163      | 1      | 816        | 409       | [gi 304155]      | [diaminopimelate decarboxylase (Bacillus methanolicus) sp p41023 DCDA_BACMT DIAMINOPIMELATE DECARBOXYLASE (EC 4.1.1.20) DAP DECARBOXYLASE]     | 85    | 62      | 408         |
| 1924      | 1      | 487        | 251       | [gi 215098]      | [lexAionase (Bacteriophage 154a)]  | 85    | 73      | 237         |
| 2912      | 1      | 776        | 390       | [gi 1041099]     | [Pyruvate Kinase (Bacillus licheniformis)]   | 85    | 71      | 387         |
| 3010      | 1      | 3          | 275       | [gi 42370]       | [pyruvate formate-lyase (AA 1-760) (Escherichia coli) ir S01788 S01788 formate C-acetyltransferase (EC 2.3.1.54) - cherichia coli]             | 85    | 74      | 273         |
| 3111      | 1      | 595        | 299       | [gi 63568]       | [limb deformity protein (Gallus gallus)]   | 85    | 85      | 297         |
| 3778      | 1      | 630        | 316       | [gi 391840]      | [beta-subunit of HDT (Pseudomonas fragi)]  | 85    | 67      | 315         |
| 3835      | 1      | 1          | 387       | [gi 1204472]     | [type I restriction enzyme ECO124/3 I M protein (Haemophilus influenzae)]  | 85    | 56      | 387         |
| 4042      | 1      | 3          | 386       | [gi 118178]      | [formate acetyltransferase (Chlamydomonas reinhardtii) ir S24997 S24997 formate C-acetyltransferase (EC 2.3.1.54) - Chlamydomonas reinhardtii] | 85    | 70      | 384         |
| 4053      | 1      | 35         | 340       | [gi 1204472]     | [type I restriction enzyme ECO124/3 I M protein (Haemophilus influenzae)]  | 85    | 56      | 306         |
| 4108      | 1      | 2          | 181       | [gi 1072418]     | [glcA gene product (Staphylococcus carnosus)]  | 85    | 61      | 180         |
| 4300      | 1      | 575        | 330       | [gi 1151932]     | [fructose enzyme II (Rhodobacter capsulatus)]  | 85    | 59      | 246         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 4392      | 1      | 627        | 355       | gi1022725       | unknown [Staphylococcus haemolyticus]   | 85    | 74      | 273         |
| 4408      | 1      | 2          | 235       | gi1871784       | Cip-like ATP-dependent protease binding subunit [Bos taurus]  | 85    | 62      | 234         |
| 4430      | 1      | 578        | 291       | gi1005166       | Respiratory nitrate reductase [Bacillus subtilis]   | 85    | 68      | 288         |
| 4555      | 1      | 2          | 253       | gi1450888       | hadh gene of Ecopri gene product [Escherichia coli] pir[S38437/S38437 hadh protein - Escherichia coli pir/S09629/S09629 hypothetical protein A - Escherichia coli (S08 40-520)] | 85    | 52      | 252         |
| 4611      | 1      | 481        | 242       | gi1256635       | dihydroxy-acid dehydratase [Bacillus subtilis]  | 85    | 65      | 240         |
| 4         | 10     | 10861      | 10591     | gi146982        | foad gene product [Staphylococcus epidermidis]  | 84    | 68      | 531         |
| 13        | 2      | 1348       | 1172      | gi162450        | ehrc protein [Bacillus subtilis]  | 84    | 56      | 177         |
| 16        | 4      | 1803       | 4852      | gi1277198       | ONA repair protein [Deinococcus radiodurans]  | 84    | 67      | 2850        |
| 22        | 3      | 1535       | 1128      | gi1511069       | Uraf [Staphylococcus xylosus]   | 84    | 73      | 408         |
| 23        | 7      | 5055       | 5306      | gi1603320       | Yer082p [Saccharomyces cerevisiae]  | 84    | 61      | 252         |
| 53        | 11     | 11587      | 11145     | gi1303948       | Vqiw [Bacillus subtilis]  | 84    | 68      | 453         |
| 53        | 12     | 14059      | 12770     | gi142613        | branched chain alpha-keto acid dehydrogenase E2 [Bacillus subtilis]   | 84    | 71      | 1270        |
| 70        | 1      | 1332       | 982       | gi146647        | ORF (repE) [Staphylococcus aureus]  | 84    | 68      | 351         |
| 73        | 4      | 2512       | 4311      | gi142593        | glycerol-3-phosphate dehydrogenase (gldP) (EC 1.1.99.5) [Bacillus subtilis]   | 84    | 74      | 1800        |
| 9A        | 7      | 4324       | 6096      | gi1467427       | [methionyl]-tRNA synthetase [Bacillus subtilis]   | 84    | 66      | 1773        |
| 100       | 9      | 9501       | 8680      | gi1340128       | ORF1 [Staphylococcus aureus]  | 84    | 78      | 822         |
| 117       | 3      | 1934       | 3208      | gi1237019       | Srb [Bacillus subtilis]   | 84    | 68      | 1275        |
| 148       | 6      | 4720       | 5670      | gi1467462       | cysteine synthetase A [Bacillus subtilis]   | 84    | 69      | 931         |
| 152       | 4      | 2064       | 2456      | gi143377        | pyruvate decarboxylase (E-1) alpha subunit [Bacillus subtilis]  | 84    | 70      | 393         |
| 169       | 7      | 3634       | 3461      | gi1001342       | hypothetical protein [Synecocystis sp.]   | 84    | 66      | 228         |
| 171       | 4      | 2992       | 2857      | gi1517475       | D-amino acid transaminase [Staphylococcus haemolyticus]   | 84    | 71      | 336         |
| 186       | 6      | 6941       | 6216      | gi1467475       | unknown [Bacillus subtilis]   | 84    | 70      | 726         |
| 205       | 9      | 6261       | 5692      | gi1216340       | ORF for adenylate kinase [Bacillus subtilis]  | 84    | 71      | 570         |
| 224       | 2      | 915        | 1391      | gi1288269       | beta-fructofuranosidase [Staphylococcus xylosus]  | 84    | 70      | 477         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name   | % ident | Length (nt) |
|-----------|--------|------------|-----------|-----------------|---|---------|-------------|
| 251       | 1      | 92         | 388       | gi 1303790      | YgeI [Bacillus subtilis]  | 84      | 65          |
| 282       | 3      | 1526       | 2836      | gi 143040       | glutamate-1-semialdehyde 2,1-aminotransferase [Bacillus subtilis]<br>pir 042728 042728 glutamate-1-semialdehyde 2,1-aminotransferase [EC 4.3.8] - Bacillus subtilis | 84      | 75          |
| 307       | 5      | 1136       | 2959      | gi 1070014      | protein-dependent [Bacillus subtilis]   | 84      | 62          |
| 320       | 4      | 2243       | 4229      | gi 143390       | [carbamyl phosphate synthetase [Bacillus subtilis]  | 84      | 70          |
| 372       | 1      | 3          | 286       | gi 1022725      | unknown [Staphylococcus hemolyticus]  | 84      | 70          |
| 411       | 2      | 2201       | 3341      | gi 1256146      | YbqI [Bacillus subtilis]  | 84      | 65          |
| 439       | 1      | 3          | 392       | gi 1046173      | osmotically inducible protein [Mycoplasma genitalium]   | 84      | 53          |
| 461       | 3      | 1362       | 2270      | gi 40211        | threonine synthase (thrC) [AA 1-352] [Bacillus subtilis] ir A25364 A25364<br>threonine synthase [EC 4.2.99.2] - Bacillus subtilis                                   | 84      | 69          |
| 487       | 1      | 3          | 299       | gi 1144531      | integritin-like protein alpha Intip [Candida albicans]  | 84      | 46          |
| 491       | 2      | 624        | 905       | pir S08564 R365 | ribosomal protein S9 - Bacillus stearothermophilus  | 84      | 69          |
| 491       | 3      | 816        | 1033      | pir S08564 R365 | ribosomal protein S9 - Bacillus stearothermophilus  | 84      | 77          |
| 548       | 1      | 3          | 341       | gi 411231       | uracil permease [Bacillus caldolyticus]   | 84      | 74          |
| 728       | 2      | 2701       | 1748      | gi 912445       | DNA polymerase [Bacillus caldolyticus]  | 84      | 68          |
| 769       | 1      | 3          | 257       | gi 1510953      | cobalamin biosynthesis protein N [Methanococcus jannaschii]   | 84      | 38          |
| 954       | 1      | 308        | 196       | gi 1405454      | aconitase [Bacillus subtilis]   | 84      | 57          |
| 957       | 1      | 3          | 395       | gi 143402       | recombination protein (ttg start codon) [Bacillus subtilis] gi 1303923 RecN [Bacillus subtilis]   | 84      | 68          |
| 975       | 1      | 3          | 452       | gi 885934       | CtpA [Synecoccus sp.]   | 84      | 70          |
| 1385      | 1      | 3          | 257       | gi 1510140      | ligandopeptidase F [Lactococcus lactis]   | 84      | 56          |
| 2954      | 1      | 3          | 323       | gi 603769       | HutU protein, urecanase [Bacillus subtilis]   | 84      | 73          |
| 2996      | 1      | 650        | 348       | gi 18178        | formate acetyltransferase [Chlamydomonas reinhardtii] ir S24997 S24997<br>formate C-acetyltransferase [EC 2.3.1.54] - Chlamydomonas reinhardtii                     | 84      | 65          |
| 3766      | 1      | 717        | 375       | gi 517205       | 67 kDa Myosin-crossreactive streptococcal antigen [Streptococcus yogenes]   | 84      | 72          |
| 4032      | 1      | 2          | 169       | gi 1146206      | glutamate dehydrogenase [Bacillus subtilis]   | 84      | 54          |
| 4058      | 1      | 620        | 312       | gi 151932       | fructose enzyme II [Rhodospirillum rubrum]  | 84      | 71          |
| 4108      | 2      | 106        | 351       | gi 1072418      | glcA gene product [Staphylococcus carnosus]   | 84      | 77          |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name   | % sim | % ident | Length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 4183      | 1      | 3          | 308       | gi 603769       | lutU protein, urocanase [Bacillus subtilis]   | 84    | 72      | 306         |
| 4726      | 1      | 55         | 234       | gi 146208       | glutamate synthase large subunit (EC 2.6.1.53) [Escherichia coli]<br>pir A29617 A29617 glutamate synthase (NADPH) (EC 1.4.1.13) large chain -<br>Escherichia coli | 84    | 73      | 180         |
| 22        | 4      | 2043       | 1576      | gi 393297       | urease accessory protein [Bacillus sp.]   | 83    | 64      | 468         |
| 53        | 13     | 14722      | 13745     | gi 142612       | branched chain alpha-keto acid dehydrogenase E1-beta [Bacillus subtilis]  | 83    | 68      | 978         |
| 57        | 16     | 13357      | 12872     | gi 143132       | lactate dehydrogenase (AC 1.1.1.27) [Bacillus caldolyticus]<br>pir B29704 B29704 L-lactate dehydrogenase (EC 1.1.1.27) - Bacillus<br>caldolyticus                 | 83    | 66      | 486         |
| 66        | 3      | 3139       | 2274      | gi 1303894      | yqhM [Bacillus subtilis]  | 83    | 63      | 846         |
| 66        | 5      | 6118       | 4643      | gi 1212730      | yqhK [Bacillus subtilis]  | 83    | 68      | 1476        |
| 70        | 3      | 1864       | 1523      | gi 144095       | replication initiator protein [Listeria monocytogenes]  | 83    | 73      | 342         |
| 90        | 1      | 377        | 1429      | gi 155571       | alcohol dehydrogenase 1 (adhA) (EC 1.1.1.1) [Zymomonas mobilis]<br>pir A35260 A35260 alcohol dehydrogenase (EC 1.1.1.1) - Zymomonas oblique                       | 83    | 70      | 1053        |
| 95        | 2      | 708        | 2162      | gi 506381       | phospho-beta-glucosidase [Bacillus subtilis]  | 83    | 70      | 1455        |
| 137       | 1      | 68         | 694       | gi 467391       | initiation protein of replicon [Bacillus subtilis]  | 83    | 77      | 627         |
| 140       | 4      | 3205       | 2742      | gi 634107       | kdsB [Escherichia coli]   | 83    | 65      | 468         |
| 142       | 3      | 3468       | 2989      | gi 1212776      | lunaxine synthase (b-subunit) [Bacillus amyloliquefaciens]  | 83    | 69      | 480         |
| 161       | 12     | 5749       | 6696      | gi 903307       | ORF75 [Bacillus subtilis]   | 83    | 64      | 948         |
| 164       | 9      | 9880       | 11070     | gi 49316        | ORF2 gene product [Bacillus subtilis]   | 83    | 66      | 1191        |
| 164       | 14     | 14148      | 14546     | gi 580902       | ORF6 gene product [Bacillus subtilis]   | 83    | 60      | 399         |
| 170       | 2      | 3144       | 2467      | gi 520844       | orf4 [Bacillus subtilis]  | 83    | 64      | 678         |
| 186       | 2      | 2029       | 1370      | gi 289284       | cysteineyl-tRNA synthetase [Bacillus subtilis]  | 83    | 72      | 660         |
| 205       | 14     | 7822       | 7607      | gi 216337       | ORF for L30 ribosomal protein [Bacillus subtilis]   | 83    | 74      | 216         |
| 237       | 6      | 3683       | 4540      | gi 1510488      | limdase/acyl-CoA:phosphate synthase (cyclase) [Methanococcus jannaschii]  | 83    | 60      | 858         |
| 301       | 1      | 985        | 638       | gi 467419       | unknown [Bacillus subtilis]   | 83    | 65      | 348         |
| 302       | 4      | 1421       | 2743      | gi 508979       | GTP-binding protein [Bacillus subtilis]   | 83    | 68      | 1323        |
| 321       | 4      | 3933       | 3571      | gi 139844       | fumarate (fcdC) (aa 1-467) [Bacillus subtilis]  | 83    | 68      | 363         |
| 367       | 1      | 2          | 352       | gi 1039479      | ORFU [Lactococcus lactis]   | 83    | 54      | 331         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 387       | 1      | 3          | 662       | gi 806281       | DNA polymerase I (Bacillus stearothermophilus)   | 83    | 70      | 660         |
| 527       | 2      | 916        | 1566      | gi 396259       | protease (Staphylococcus epidermidis)  | 83    | 67      | 651         |
| 533       | 1      | 355        | 179       | gi 142455       | alanine dehydrogenase (EC 1.4.1.1) (Bacillus stearothermophilus)<br>pir b34261 b34261 alanine dehydrogenase (EC 1.4.1.1) - Bacillus stearothermophilus | 83    | 66      | 177         |
| 536       | 4      | 1617       | 1438      | gi 143366       | adenylosuccinate lyase (Pur-B) (Bacillus subtilis); r C39326 w28SDS  | 83    | 67      | 180         |
| 652       | 1      | 2          | 859       | gi 520753       | adenylosuccinate lyase (EC 4.3.2.2) - Bacillus subtilis  | 83    | 72      | 858         |
| 778       | 2      | 200        | 361       | gi 1522665      | M. jannaschii predicted coding region MJC128 (Methanococcus jannaschii)  | 83    | 58      | 182         |
| 897       | 1      | 120        | 296       | gi 1064807      | ORTHONINE AMINOTRANSFERASE (Bacillus subtilis)   | 83    | 76      | 177         |
| 1213      | 1      | 3          | 491       | gi 289288       | lexA (Bacillus subtilis)   | 83    | 67      | 489         |
| 2529      | 1      | 296        | 350       | gi 143786       | tryptophanyl-tRNA synthetase (EC 6.1.1.2) (Bacillus subtilis)<br>pir b70481 w85 tryptophan-tRNA ligase (EC 6.1.1.2) - Bacillus subtilis                | 83    | 69      | 147         |
| 2573      | 1      | 649        | 326       | gi 1109687      | Prox (Bacillus subtilis)   | 83    | 58      | 324         |
| 1009      | 1      | 728        | 366       | gi 142532       | ORF_0294 (Bacillus coli)   | 83    | 65      | 363         |
| 3035      | 2      | 45         | 305       | gi 950082       | hypothetical yeast protein 1 (Mycoplasma capricolum) pir 549578 549578<br>hypothetical protein - Mycoplasma capricolum 56C3 (fragment)                 | 83    | 59      | 261         |
| 1906      | 1      | 87         | 309       | gi 135197       | chloroform reductase (Eubacterium acidaminophilum)   | 83    | 61      | 243         |
| 445R      | 1      | 540        | 271       | gi 397526       | clumping factor (Staphylococcus aureus)  | 83    | 78      | 270         |
| 4570      | 1      | 444        | 223       | gi 1022726      | unknown (Staphylococcus haemolyticus)  | 83    | 74      | 222         |
| 4654      | 1      | 97         | 261       | gi 1072419      | glcB gene product (Staphylococcus carnosus)  | 83    | 79      | 165         |
| 16        | 2      | 295        | 1191      | gi 153854       | lvs402 protein (Streptococcus pneumoniae)  | 82    | 67      | 897         |
| 16        | 3      | 1193       | 1798      | gi 153854       | lvs402 protein (Streptococcus pneumoniae)  | 82    | 70      | 606         |
| 38        | 12     | 9644       | 8724      | gi 1204400      | N-acetylneuraminase lyase (Haemophilus influenzae)   | 82    | 56      | 923         |
| 42        | 4      | 988        | 2019      | gi 1841192      | catalase (Bacteroides fragilis)  | 82    | 70      | 1032        |
| 51        | 6      | 2590       | 3489      | gi 143607       | sporulation protein (Bacillus subtilis)  | 82    | 69      | 900         |
| 56        | 11     | 12270      | 13925     | gi 39431        | oligo-1,6-glucosidase (Bacillus cereus)  | 82    | 60      | 1656        |
| 56        | 15     | 17673      | 18014     | gi 467410       | unknown (Bacillus subtilis)  | 82    | 66      | 342         |
| 61        | 2      | 881        | 3313      | gi 143148       | transfer RNA-Leu synthetase (Bacillus subtilis)  | 82    | 70      | 2433        |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 82        | 7      | 9162       | 11318     | gi 48240        | elongation factor G (AA 1-691) [Thermus aquaticus thermophilus]<br> r S15928 EFTW elongation factor G - Thermus aquaticus<br> p P13551 EFC_THETH ELONGATION FACTOR G (EF-G)  | 82    | 64      | 2157        |
| 85        | 2      | 5470       | 3260      | gi 143369       | phosphoribosylformyl glycylamide synthetase II (PUR-Q) [Bacillus subtilis]   | 82    | 66      | 2211        |
| 102       | 6      | 3662       | 5380      | gi 1256635      | dihydroxy-acid dehydratase [Bacillus subtilis]   | 82    | 65      | 1719        |
| 117       | 4      | 3242       | 3493      | pir A47154 A471 | orf1 5' of Ffh - Bacillus subtilis   | 82    | 53      | 252         |
| 128       | 6      | 4377       | 5933      | gi 460258       | phosphoglycerate mutase [Bacillus subtilis]  | 82    | 66      | 1557        |
| 129       | 2      | 1229       | 2182      | gi 403373       | glycerophosphoryl diester phosphodiesterase [Bacillus subtilis]<br> pir S37251 S37251 glycerophosphoryl diester phosphodiesterase - acillus<br>subtilis  | 82    | 62      | 954         |
| 170       | 1      | 2          | 1441      | gi 1377831      | unknown [Bacillus subtilis]  | 82    | 67      | 1440        |
| 177       | 1      | 3          | 1094      | gi 467386       | thiophen and furan oxidation [Bacillus subtilis]   | 82    | 65      | 1092        |
| 184       | 4      | 3572       | 4039      | gi 153566       | ORF (19K protein) [Enterococcus faecalis]  | 82    | 59      | 468         |
| 189       | 8      | 4455       | 4225      | gi 1001878      | CspL protein [Listeria monocytogenes]  | 82    | 73      | 231         |
| 206       | 19     | 21366      | 20707     | gi 473916       | lipopeptide antibiotic (curin A) [Bacillus subtilis] sp P39144 P14_BACSU<br>LIPOPEPTIDE ANTIBIOTICS ITURIN A AND SURFACTIN BIOSYNTHESIS PROTEIN  | 82    | 50      | 660         |
| 221       | 2      | 805        | 1722      | gi 517205       | 67 kDa Myosin-crossreactive streptococcal antigen [Streptococcus pyogenes]   | 82    | 63      | 918         |
| 223       | 4      | 3866       | 3651      | gi 439619       | [Salmonella typhimurium] IS200 insertion sequence from SRA17, artificial<br>gene product [Salmonella typhimurium]  | 82    | 69      | 216         |
| 260       | 3      | 5207       | 4296      | gi 1161381      | IcA8 [Staphylococcus epidermidis]  | 82    | 61      | 912         |
| 315       | 3      | 4864       | 2855      | gi 143397       | quinol oxidase [Bacillus subtilis]   | 82    | 67      | 2010        |
| 321       | 10     | 8520       | 7945      | gi 143981       | ORF5: This ORF includes a region (aa23-103) containing a potential non-sulphur centre homologous to a region of Rhodospirillum rubrum and Chromatium vinosum; putative [Bacillus stearothermophilus]<br> pir P02029 P02029 hypothetical protein 5 (gldA 3' region) - | 82    | 62      | 576         |
| 331       | 3      | 1055       | 1342      | gi 436574       | ribosomal protein L1 [Bacillus subtilis]   | 82    | 71      | 288         |
| 370       | 2      | 262        | 618       | gi 1303793      | Vqel [Bacillus subtilis]   | 82    | 59      | 357         |
| 404       | 4      | 3053       | 4024      | gi 1303821      | Vqte [Bacillus subtilis]   | 82    | 68      | 972         |
| 405       | 4      | 4440       | 3073      | gi 1303913      | Vqhx [Bacillus subtilis]   | 82    | 67      | 1168        |
| 436       | 3      | 4096       | 2864      | gi 149521       | cryptophan synthase beta subunit [Lactococcus lactis]  pir S35129 S35129<br>cryptophan synthase (EC 4.2.1.20) beta chain - actococcus lactis subsp.<br>lactis  | 82    | 67      | 1233        |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 441       | 4      | 3394       | 2573      | gi1142952       | glyceroldehyde-3-phosphate dehydrogenase [Bacillus teurothermophilus]   | 82    | 67      | 822         |
| 444       | 12     | 10415      | 11222     | gi11204354      | spore germination and vegetative growth protein [Haemophilus influenzae]  | 82    | 67      | 813         |
| 446       | 1      | 3          | 191       | gi1143387       | aspartate transcarbamylase [Bacillus subtilis]  | 82    | 66      | 189         |
| 462       | 3      | 1007       | 1210      | gi1142521       | deoxyribodipyridine photolysis [Bacillus subtilis] p1c[A37192][A37192 UVIR protein - Bacillus subtilis sp14951][UVR-C-BACSU EXCINUCLEASE ABC SUBUNIT C]   | 82    | 64      | 204         |
| 537       | 1      | 1560       | 784       | gi1853767       | UDP-N-acetylglucosamine 1-carboxyvinyltransferase [Bacillus subtilis]   | 82    | 61      | 777         |
| 680       | 2      | 407        | 700       | gi1426472       | secE gene product [Staphylococcus carnosus]   | 82    | 69      | 294         |
| 724       | 2      | 585        | 386       | gi1143373       | phosphoribosyl aminimidazole carboxy formyl ormyltransferase/inosine monophosphate cyclohydrolase (PUN-H(D)) [Bacillus subtilis]  | 82    | 68      | 180         |
| 763       | 1      | 422        | 213       | gi1467458       | cell division protein [Bacillus subtilis]   | 82    | 35      | 210         |
| 818       | 1      | 584        | 281       | gi1064787       | function unknown [Bacillus subtilis]  | 82    | 69      | 282         |
| 858       | 1      | 175        | 1176      | gi1143043       | uroporphyrinogen decarboxylase [Bacillus subtilis] p1c[B47045][B47045 uroporphyrinogen decarboxylase (EC 4.1.1.37) - acillus subtilis]  | 82    | 71      | 1002        |
| 895       | 1      | 3          | 599       | gi1027507       | ATP binding protein [Borrelia burgdorferi]  | 82    | 72      | 597         |
| 939       | 1      | 10         | 399       | gi1143795       | transfer RNA-Tyr synthetase [Bacillus subtilis]   | 82    | 60      | 390         |
| 961       | 1      | 3          | 306       | gi1577647       | gamma-hemolysin [Staphylococcus aureus]   | 82    | 69      | 306         |
| 1192      | 1      | 307        | 155       | gi1146974       | NH3-dependent NAD synthetase [Escherichia coli]   | 82    | 71      | 153         |
| 1317      | 1      | 49         | 375       | gi1407908       | Elisact [Staphylococcus xylosum]  | 82    | 72      | 327         |
| 1341      | 1      | 1          | 150       | gi139962        | ribosomal protein L35 (AA 1-66) [Bacillus stearothermophilus] ir[S05347][R58535 ribosomal protein L35 - Bacillus stearothermophilus]  | 82    | 68      | 150         |
| 2990      | 2      | 567        | 349       | gi1534855       | ATPase subunit epsilon [Bacillus stearothermophilus] sp142009[ATPE-BACST ATP SYNTHASE EPSILON CHAIN (EC 3.6.1.34)]  | 82    | 47      | 219         |
| 3024      | 1      | 45         | 224       | gi1467402       | unknown [Bacillus subtilis]   | 82    | 64      | 180         |
| 3045      | 1      | 276        | 139       | gi1467335       | ribosomal protein L9 [Bacillus subtilis]  | 82    | 60      | 138         |
| 3045      | 2      | 558        | 400       | gi1467335       | ribosomal protein L9 [Bacillus subtilis]  | 82    | 82      | 159         |
| 3091      | 1      | 474        | 238       | gi1499335       | secA protein [Staphylococcus carnosus]  | 82    | 78      | 237         |
| 3107      | 1      | 416        | 210       | gi1546918       | orfY 3' of comK [Bacillus subtilis, E26, Peptide Partial, 140 aa] p1c[S4312][S43612 hypothetical protein Y - Bacillus subtilis sp140358][VHKD-BACSU HYPOTHETICAL PROTEIN IN COMK 3'-REGION (ORF7) FRAGMENT] | 82    | 64      | 207         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 432       | 1      | 2          | 319       | gi142086        | nitrate reductase alpha subunit [Escherichia coli] p[P09152]NARO_ECOLI<br>RESPIRATORY NITRATE REDUCTASE 1 ALPHA CHAIN (EC 7.99.4). (SUB 2-1247)                              | 82    | 75      | 318         |
| 23        | 3      | 3275       | 2574      | gi1199573       | sp08 [Sphingomonas sp.]  | 81    | 64      | 702         |
| 42        | 1      | 638        | 321       | gi1466778       | lysine specific peptidase [Escherichia coli]   | 81    | 59      | 318         |
| 48        | 5      | 4051       | 4350      | gi11045937      | M. genitalium predicted coding region NG246 [Mycoplasma genitalium]  | 81    | 62      | 300         |
| 51        | 4      | 1578       | 2579      | gi1516649       | dcIAC protein - Bacillus subtilis  | 81    | 55      | 1002        |
| 53        | 2      | 354        | 1494      | gi11303961      | YqjJ [Bacillus subtilis]   | 81    | 67      | 1131        |
| 53        | 8      | 9419       | 7971      | gi1146930       | 6-phosphogluconate dehydrogenase [Escherichia coli]  | 81    | 66      | 1449        |
| 54        | 9      | 10757      | 10119     | gi1143016       | permease [Bacillus subtilis]   | 81    | 65      | 639         |
| 54        | 10     | 13360      | 11786     | gi1143015       | gluconate kinase [Bacillus subtilis]   | 81    | 64      | 1575        |
| 57        | 17     | 13983      | 13366     | gi1425605       | L-lactate dehydrogenase (EC 1.1.1.27) - Bacillus subtilis  | 81    | 74      | 618         |
| 81        | 2      | 2708       | 2217      | gi1122302       | NiG-related protein [Haemophilus influenzae]   | 81    | 54      | 492         |
| 86        | 1      | 745        | 374       | gi1414017       | /jpa-93D gene product [Bacillus subtilis]  | 81    | 70      | 372         |
| 103       | 6      | 6438       | 4861      | gi11971342      | nitrate reductase beta subunit [Bacillus subtilis] sp[P42176]NARO_BACSU<br>NITRATE REDUCTASE BETA CHAIN (EC 1.7.99.4).   | 81    | 64      | 1578        |
| 120       | 15     | 110845     | 12338     | gi11524392      | GbsA [Bacillus subtilis]   | 81    | 67      | 1494        |
| 128       | 5      | 3676       | 4413      | gi1143319       | triose phosphate isomerase [Bacillus megaterium]   | 81    | 64      | 738         |
| 131       | 9      | 110308     | 9280      | gi1299163       | alanine dehydrogenase [Bacillus subtilis]  | 81    | 68      | 1029        |
| 143       | 6      | 6088       | 5471      | gi1439619       | [Salmonella typhimurium] IS700 insertion sequence from SAR17, att1a1.1.<br>gene product [Salmonella typhimurium]   | 81    | 61      | 618         |
| 169       | 1      | 43         | 825       | gi1897795       | 30S ribosomal protein (pediococcus acidilactici) sp[P49668]RS2_PEDAC 30S<br>RIBOSOMAL PROTEIN S2   | 81    | 65      | 783         |
| 230       | 1      | 450        | 226       | gi1125826       | short region of weak similarity to tyrosine-protein kinase receptors in a<br>fibronectin type III-like domain [Caenorhabditis elegans]                                       | 81    | 54      | 225         |
| 233       | 5      | 2000       | 2677      | gi1467404       | unknown [Bacillus subtilis]  | 81    | 63      | 678         |
| 241       | 2      | 3081       | 2149      | gi116510        | succinate-CoA ligase (GDP-forming) [Arabidopsis thaliana] ir[S30579]S30579<br>succinate-CoA ligase (GDP-forming) (EC 6.2.1.4) pha chain - Arabidopsis<br>thaliana (fragment) | 81    | 69      | 913         |
| 256       | 1      | 1          | 981       | gi1509411       | spolIIC protein - Bacillus subtilis  | 81    | 65      | 981         |
| 259       | 3      | 3752       | 2693      | gi128167        | PROBABLE PEPTIDE CHAIN RELEASE FACTOR 2 (RF-2) (FRAGMENT).   | 81    | 65      | 1062        |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 275       | 2      | 1728       | 3581      | gi1726480       | L-glutamine-D-fructose-6-phosphate amidotransferase (Bacillus subtilis)  | 81    | 68      | 1854        |
| 285       | 1      | 1466       | 735       | gi1204844       | H. influenzae predicted coding region H10594 (Haemophilus influenzae)  | 81    | 63      | 732         |
| 296       | 1      | 99         | 1406      | gi1467328       | adenylosuccinate synthetase (Bacillus subtilis)  | 81    | 67      | 1108        |
| 302       | 9      | 5590       | 5889      | gi147485        | queA (Escherichia coli)  | 81    | 64      | 300         |
| 317       | 2      | 1137       | 1376      | gi134961        | resolvase (Transposon Tn917)   | 81    | 57      | 240         |
| 343       | 2      | 1034       | 1342      | gi1405955       | lyed (Escherichia coli)  | 81    | 60      | 309         |
| 360       | 2      | 1404       | 2471      | gi1204570       | aspartyl-tRNA synthetase (Haemophilus influenzae)  | 81    | 67      | 1068        |
| 364       | 5      | 6251       | 5706      | gi1204652       | methylated-DNA--protein-cysteine methyltransferase (Haemophilus influenzae)  | 81    | 63      | 546         |
| 372       | 2      | 1707       | 1135      | gi1467416       | unknown (Bacillus subtilis)  | 81    | 65      | 573         |
| 392       | 1      | 43         | 603       | pir1509411 5094 | spor11E protein - Bacillus subtilis  | 81    | 65      | 561         |
| 404       | 9      | 5252       | 6154      | gi1606745       | Bex (Bacillus subtilis)  | 81    | 65      | 903         |
| 426       | 2      | 1727       | 1119      | gi139453        | Manganese superoxide dismutase (Bacillus caldotenax) Ir[S22053]S22053 superoxide dismutase (EC 1.15.1.1) (Mn) - Bacillus licheniformis | 81    | 66      | 609         |
| 440       | 7      | 5653       | 5889      | pir1537081 C370 | hypothetical protein 31 (contig 3, region) - Salmonella typhimurium (fragment)   | 81    | 57      | 237         |
| 625       | 3      | 1105       | 2070      | gi11262360      | protein kinase PknB (Mycobacterium leprae)   | 81    | 56      | 966         |
| 754       | 2      | 504        | 1064      | gi11303902      | YqjU (Bacillus subtilis)   | 81    | 71      | 561         |
| 842       | 1      | 86         | 430       | gi11405446      | transketolase (Bacillus subtilis)  | 81    | 68      | 345         |
| 953       | 1      | 798        | 400       | gi11205429      | dipeptide transport ATP-binding protein (Haemophilus influenzae)   | 81    | 57      | 399         |
| 961       | 2      | 252        | 401       | gi1487686       | synergohymenotrophic toxin (Staphylococcus intermedius) pIr[S4494]S4494 synergohymenotrophic toxin - Staphylococcus intermedius        | 81    | 72      | 150         |
| 1035      | 1      | 1          | 189       | gi11046138      | M. genitalium predicted coding region MG523 (Mycoplasma genitalium)  | 81    | 43      | 189         |
| 1280      | 1      | 670        | 449       | gi1559164       | helicase (AutoGrapha californica nuclear polyhedrosis virus) ap1P24307 v143.NPVAC HELICASE   | 81    | 43      | 222         |
| 3171      | 1      | 68         | 241       | gi1322245       | mevalonate pyrophosphate decarboxylase (Rattus norvegicus)   | 81    | 62      | 174         |
| 3715      | 1      | 475        | 239       | gi1537137       | ORF_138 (Escherichia coli)   | 81    | 58      | 237         |
| 3908      | 1      | 2          | 325       | gi1439619       | Salmonella typhimurium IS200 insertion sequence from SAR17, attial.1, gene product (Salmonella typhimurium)                            | 81    | 68      | 324         |
| 3940      | 1      | 3          | 401       | pir1296464      | ATPase (Lactococcus lactis)  | 81    | 69      | 399         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match      | accession | match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|------------|-----------|--|-------|---------|-------------|
| 3954      | 1      | 1          | 318       | gi 1224069 |           | amidase [Morsakella catarrhalis]                                   | 81    | 68      | 318         |
| 4049      | 1      | 337        | 170       | gi 603768  |           | MutI protein, imidazole-5-propionate hydrolase [Bacillus subtilis] | 81    | 68      | 168         |
| 4209      | 1      | 1          | 324       | gi 403373  |           | glycerophosphoryl diester phosphodiesterase [Bacillus subtilis]    | 81    | 58      | 324         |
| 4371      | 1      | 627        | 322       | gi 216677  |           | indolepyruvate decarboxylase [Enterobacter cloacae] pir[S16013]    | 81    | 72      | 306         |
| 4387      | 1      | 19         | 228       | gi 460689  |           | indolepyruvate decarboxylase (EC 4.1.1.-) - enterobacter cloacae   | 81    | 59      | 210         |
| 4191      | 1      | 581        | 306       | gi 1524193 |           | TWG [Thermactinomyces vulgaris]                                    | 81    | 57      | 276         |
| 4425      | 1      | 3          | 341       | gi 143015  |           | unknown [Mycobacterium tuberculosis]                               | 81    | 66      | 339         |
| 9         | 1      | 1593       | 847       | gi 1064786 |           | glucanase kinase [Bacillus subtilis]                               | 80    | 62      | 747         |
| 17        | 1      | 544        | 311       | gi 559164  |           | function unknown [Bacillus subtilis]                               | 80    | 40      | 234         |
| 45        | 2      | 1159       | 2448      | gi 109684  |           | halicase [Autographa californica nuclear polyhedrosis virus]       | 80    | 63      | 1290        |
| 45        | 5      | 4032       | 4733      | gi 1109687 |           | epiP24307/V143_MPVAC_HELICASE                                      | 80    | 55      | 702         |
| 54        | 8      | 10266      | 9502      | gi 563952  |           | proV [Bacillus subtilis]   | 80    | 82      | 765         |
| 62        | 12     | 8852       | 7545      | gi 854655  |           | proZ [Bacillus subtilis]   | 80    | 82      | 1308        |
| 62        | 14     | 8087       | 8683      | gi 559713  |           | glucanase permease [Bacillus licheniformis]                        | 80    | 68      | 597         |
| 67        | 16     | 13781      | 14122     | gi 305002  |           | Na/H antiporter system [Bacillus alcalophilus]                     | 80    | 65      | 342         |
| 70        | 13     | 11495      | 10296     | gi 303395  |           | ONF [Homo sapiens]   | 80    | 64      | 1200        |
| 98        | 9      | 6336       | 7130      | gi 467428  |           | ONF_E336 [Escherichia coli]  | 80    | 68      | 795         |
| 98        | 10     | 7294       | 7833      | gi 467430  |           | YqkM [Bacillus subtilis]   | 80    | 64      | 540         |
| 98        | 11     | 7820       | 8737      | gi 467431  |           | unknown [Bacillus subtilis]  | 80    | 61      | 918         |
| 109       | 16     | 14154      | 14813     | gi 580875  |           | high level kanamycin resistance [Bacillus subtilis]                | 80    | 63      | 660         |
| 112       | 15     | 14294      | 16636     | gi 1072361 |           | ipa-57d gene product [Bacillus subtilis]                           | 80    | 65      | 2343        |
| 139       | 1      | 1448       | 726       | gi 506699  |           | pyruvate-formate-lyase [Clostridium pasteurianum]                  | 80    | 58      | 723         |
| 139       | 2      | 2179       | 1448      | gi 506698  |           | capC [Staphylococcus aureus]                                       | 80    | 59      | 732         |
| 174       | 4      | 3271       | 2870      | gi 1116242 |           | capB [Staphylococcus aureus]                                       | 80    | 61      | 402         |
|           |        |            |           |            |           | aspartate 1-decarboxylase [Bacillus subtilis]                      |       |         |             |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 177       | 3      | 2102       | 2842      | gi1467185       | unknown [Bacillus subtilis]   | 80    | 70      | 741         |
| 184       | 6      | 6124       | 5912      | gi1161953       | 85-kDa surface antigen [Trypanosoma cruzi]  | 80    | 46      | 213         |
| 186       | 4      | 5388       | 3875      | gi1289282       | glucanase [Bacillus subtilis]   | 80    | 55      | 1494        |
| 205       | 10     | 15796      | 15140     | gi140103        | ribosomal protein L4 [Bacillus stearothermophilus]  | 80    | 56      | 657         |
| 207       | 1      | 140        | 1315      | gi1460259       | enolase [Bacillus subtilis]   | 80    | 67      | 1176        |
| 211       | 3      | 1078       | 1590      | gi1410131       | ORFX [Bacillus subtilis]  | 80    | 61      | 513         |
| 235       | 2      | 1962       | 2255      | gi1143797       | valyl-tRNA synthetase [Bacillus stearothermophilus] sp119311SV_BACST<br>VALYL-TRNA SYNTHETASE (EC 6.1.1.9) VALINE--TRNA LIGASE [VALYST]                 | 80    | 55      | 294         |
| 239       | 1      | 1          | 1263      | gi1143000       | proton glutamate symport protein [Bacillus stearothermophilus]<br>pir15262471526247 glutamate/aspartate transport protein - Bacillus stearothermophilus | 80    | 59      | 1263        |
| 272       | 5      | 2724       | 2461      | gi1709993       | hypothetical protein [Bacillus subtilis]  | 80    | 54      | 264         |
| 301       | 3      | 1446       | 1111      | gi1467418       | unknown [Bacillus subtilis]   | 80    | 58      | 336         |
| 310       | 4      | 3697       | 4501      | gi1177686       | ecuc gene product [Staphylococcus xylosus]  | 80    | 67      | 1197        |
| 310       | 6      | 5258       | 7006      | gi1348053       | acetyl-CoA synthetase [Bacillus subtilis]   | 80    | 67      | 1749        |
| 310       | 7      | 7410       | 9113      | gi1103865       | formyl-tetrahydrofolate synthetase [Streptococcus mutans]   | 80    | 67      | 1704        |
| 325       | 3      | 1114       | 1289      | gi1310325       | outer capsid protein [Rotavirus sp.]  | 80    | 40      | 276         |
| 337       | 1      | 1268       | 636       | gi1537049       | ORF_0470 [Escherichia coli]   | 80    | 55      | 633         |
| 374       | 2      | 929        | 1228      | gi1405448       | YnfF [Bacillus subtilis]  | 80    | 70      | 300         |
| 375       | 5      | 3062       | 3331      | gi1467448       | unknown [Bacillus subtilis]   | 80    | 68      | 270         |
| 388       | 1      | 267        | 587       | gi1064791       | [function unknown [Bacillus subtilis]]  | 80    | 65      | 321         |
| 394       | 1      | 9          | 659       | gi1304976       | [matches PS00017: ATP-GTP_A and PS00301: EFACITOR_GTP; similar to longation factor G, TetM/TetO tetracycline-resistance proteins Escherichia coli]      | 80    | 65      | 651         |
| 456       | 1      | 625        | 1263      | gi1146183       | [putative [Bacillus subtilis]]  | 80    | 65      | 639         |
| 475       | 1      | 1          | 654       | gi1288269       | Beta-fructofuranosidase [Staphylococcus xylosus]  | 80    | 66      | 654         |
| 544       | 2      | 1449       | 2240      | gi1529754       | IspeC [Streptococcus pyogenes]  | 80    | 50      | 792         |
| 632       | 4      | 1623       | 1871      | gi11483545      | unknown [Mycobacterium tuberculosis]  | 80    | 65      | 249         |
| 719       | 1      | 1          | 1257      | gi1064791       | [function unknown [Bacillus subtilis]]  | 80    | 68      | 1257        |
| 739       | 1      | 107        | 838       | gi1666983       | [putative ATP binding subunit [Bacillus subtilis]]  | 80    | 61      | 732         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 745       | 2      | 581        | 614       | gi 1511600      | coenzyme PQQ synthetase protein III [Methanococcus jannaschii]  | 80    | 61      | 168         |
| 822       | 1      | 17         | 679       | gi 410141       | ORF17 [Bacillus subtilis]   | 80    | 68      | 663         |
| 827       | 2      | 591        | 816       | gi 1205301      | leukotoxin secretion ATP-binding protein [Haemophilus influenzae]   | 80    | 54      | 156         |
| 1044      | 1      | 3          | 149       | gi 60632        | Vp2 [Marburg virus]   | 80    | 55      | 147         |
| 1220      | 2      | 571        | 613       | gi A61072 EP50  | gallicidarin precursor - Staphylococcus gallinarum  | 80    | 70      | 159         |
| 2519      | 1      | 75         | 275       | gi 147556       | dps [Escherichia coli]  | 80    | 45      | 201         |
| 2947      | 1      | 503        | 279       | gi 1184680      | polynucleotide phosphorylase [Bacillus subtilis]  | 80    | 62      | 225         |
| 3120      | 1      | 2          | 226       | gi 517205       | 67 kDa Myosin-crossreactive streptococcal antigen [Streptococcus pyogenes]  | 80    | 65      | 225         |
| 3191      | 1      | 294        | 148       | gi 151259       | HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas mvalonii] pir A44756 A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.          | 80    | 59      | 147         |
| 3560      | 2      | 285        | 434       | gi 217130       | photosystem I core protein B [Synecococcus vulcanus]  | 80    | 70      | 150         |
| 3655      | 1      | 47         | 346       | gi 415855       | deoxyribose aldolase [Hycoplasma hominis]   | 80    | 56      | 300         |
| 3658      | 2      | 324        | 584       | gi 551531       | 2-nitropropene dioxygenase [Williopsis saturnus]  | 80    | 54      | 261         |
| 3769      | 1      | 798        | 400       | gi 1339950      | large subunit of NADH-dependent glucanase synthase [Plectonema boryanum]  | 80    | 68      | 399         |
| 3781      | 1      | 692        | 348       | gi 166412       | NADH-glucanase synthase [Medicago sativa]   | 80    | 62      | 345         |
| 3988      | 1      | 48         | 287       | gi 1204696      | fructose-permease IIBC component [Haemophilus influenzae]   | 80    | 69      | 240         |
| 4030      | 1      | 571        | 287       | gi 1009366      | respiratory nitrate reductase [Bacillus subtilis]   | 80    | 60      | 285         |
| 4092      | 1      | 547        | 275       | gi 1370207      | orf6 [Lactobacillus sakei]  | 80    | 69      | 273         |
| 4103      | 1      | 680        | 342       | gi 39956        | IIIGc [Bacillus subtilis]   | 80    | 65      | 339         |
| 4231      | 1      | 692        | 348       | gi 289287       | UDP-glucose pyrophosphorylase [Bacillus subtilis]   | 80    | 65      | 345         |
| 4265      | 1      | 595        | 299       | gi 603768       | HutI protein, imidazolone-5-propionate hydrolase [Bacillus subtilis] gi 603768 HutI protein, imidazolone-5-propionate hydrolase [Bacillus subtilis] | 80    | 63      | 297         |
| 4504      | 1      | 498        | 250       | gi 1339950      | large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]  | 80    | 68      | 249         |
| 2         | 6      | 5998       | 6798      | gi 535351       | CodY [Bacillus subtilis]  | 79    | 63      | 801         |
| 4         | 7      | 8295       | 7057      | gi 603768       | HutI protein, imidazolone-5-propionate hydrolase [Bacillus subtilis] gi 603768 HutI protein, imidazolone-5-propionate hydrolase [Bacillus subtilis] | 79    | 64      | 1245        |
| 25        | 6      | 5273       | 5515      | pir A36728 A367 | acyl carrier protein - Rhizobium meliloti   | 79    | 65      | 243         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 59        | 2      | 1173       | 1424      | gi 147923       | threonine dehydratase 2 (EC 4.2.1.16) [Escherichia coli]   | 79    | 75      | 252         |
| 60        | 1      | 1          | 204       | gi 666115       | orf1 upstream of glucose kinase [Staphylococcus xylosum] pir[S5235][S5235] hypothetical protein 1 - Staphylococcus xylosum   | 79    | 60      | 204         |
| 81        | 1      | 3002       | 3590      | gi 466882       | ppa1; s1696_C2_189 Mycobacterium leprae  | 79    | 64      | 1413        |
| 85        | 7      | 7023       | 6905      | gi 143364       | phosphoribosyl aminimidazole carboxylase 1 (Pur-E) [Bacillus subtilis]   | 79    | 60      | 519         |
| 89        | 6      | 5660       | 4554      | gi 144906       | product homologous to E.coli thiorodoxin reductase; J.Biol.Chem. 1988) 263:9015-9019, and to P2a protein of alkyl hydroperoxide oxidase from S.Cyphomarinum; J.Biol.Chem. 11990) 265:10535-10540; pen reading frame A [Clostridium pasteurianum] | 79    | 35      | 1107        |
| 102       | 11     | 7489       | 8571      | gi 143093       | ketol-acid reductoisomerase [Bacillus subtilis] sp[P3725][LVC_BACSU KETOL-ACID REDUCTOISOMERASE (EC 1.1.1.86) ACETOHYDROXY-ACID ISOMEROREDUCTASE] (ALPHA-KETO-BETA-HYDROXYLACIL DUCTOISOMERASE)  | 79    | 64      | 1083        |
| 102       | 14     | 11190      | 12563     | gi 149428       | putative [Lactococcus lactis]  | 79    | 65      | 1374        |
| 127       | 9      | 7792       | 9372      | gi 458688       | PrcC/NF3 [Dichelobacter nodosus]   | 79    | 68      | 1581        |
| 139       | 3      | 2540       | 3983      | gi 506497       | CapA [Staphylococcus aureus]   | 79    | 55      | 558         |
| 144       | 2      | 1644       | 1156      | gi 149296       | peptide methionine sulfoxide reductase [Streptococcus pneumoniae]  | 79    | 47      | 489         |
| 148       | 2      | 529        | 1098      | gi 467437       | hypoxanthine-guanine phosphoribosyltransferase [Bacillus subtilis] gi 467437 hypoxanthine-guanine phosphoribosyltransferase [Bacillus subtilis]  | 79    | 59      | 570         |
| 150       | 1      | 965        | 591       | gi 755602       | unknown [Bacillus subtilis]  | 79    | 61      | 375         |
| 176       | 1      | 1029       | 587       | gi 297874       | fructose-bisphosphate aldolase [Staphylococcus carnosus] pir[A4994][A4994] fructose-bisphosphate aldolase (EC 4.1.2.13) - taphylococcus carnosus (strain TH300)  | 79    | 65      | 453         |
| 186       | 7      | 7584       | 6874      | gi 1314298      | ORF3: putative Ssa protein; similar to Ssa proteins from Hemophilus influenzae and Escherichia coli [Listeria monocytogenes]   | 79    | 64      | 711         |
| 205       | 16     | 8887       | 8498      | gi 1044980      | ribosomal protein L18 [Bacillus subtilis]  | 79    | 70      | 390         |
| 211       | 1      | 1          | 519       | gi 1303994      | YqkH [Bacillus subtilis]   | 79    | 62      | 519         |
| 223       | 2      | 4183       | 2801      | gi 488430       | alcohol dehydrogenase 2 [Entamoeba histolytica]  | 79    | 60      | 1381        |
| 243       | 8      | 8915       | 7896      | gi 580883       | lpa-8d gene product [Bacillus subtilis]  | 79    | 60      | 1020        |
| 279       | 4      | 3721       | 4329      | gi 413930       | lpa-6d gene product [Bacillus subtilis]  | 79    | 59      | 609         |
| 300       | 1      | 11         | 1393      | gi 403372       | glycerol 3-phosphate perase [Bacillus subtilis]  | 79    | 62      | 1383        |
| 307       | 3      | 2930       | 1935      | gi 950062       | hypothetical yeast protein 1 [Mycoplasma capricolum] pir[S48578][S48578] hypothetical protein - Mycoplasma capricolum SGC3 (fragment)  | 79    | 60      | 996         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 332       | 6      | 10106      | 8886      | gi 216834       | p47k [Pseudomonas chlororaphis]  | 79    | 59      | 1221        |
| 412       | 1      | 1153       | 578       | gi 143177       | putative [Bacillus subtilis]   | 79    | 51      | 576         |
| 481       | 3      | 621        | 1124      | gi 786163       | Ribosomal Protein L10 [Bacillus subtilis]  | 79    | 66      | 504         |
| 516       | 1      | 702        | 352       | gi 805090       | NisP [Lactococcus lactis]  | 79    | 48      | 351         |
| 525       | 2      | 2457       | 1426      | gi 143371       | phosphoribosyl aminimidazole synthetase (pur-M) [Bacillus subtilis]<br>pirH49326[AJASCU phosphoribosylformylglycinamide cyclo-ligase EC 6.3.3.1] - Bacillus subtilis | 79    | 61      | 1032        |
| 53A       | 4      | 3448       | 2825      | gi 1370207      | lorf6 [Lactobacillus sakei]  | 79    | 67      | 624         |
| 570       | 1      | 2          | 421       | gi 476160       | arginine permease substrate-binding subunit [Listeria monocytogenes]   | 79    | 61      | 420         |
| 645       | 6      | 2663       | 3241      | gi 153898       | transport protein [Salmonella typhimurium]   | 79    | 62      | 579         |
| 683       | 1      | 75         | 374       | gi 1064795      | [function unknown] [Bacillus subtilis]   | 79    | 62      | 300         |
| 816       | 3      | 4700       | 3987      | gi 1407784      | lorf-1, novel antigen [Stenohylococcus aureus]   | 79    | 62      | 714         |
| 2829      | 1      | 3          | 401       | gi 1524397      | glycine betaine transporter Opud [Bacillus subtilis]   | 79    | 61      | 399         |
| 2837      | 1      | 357        | 202       | pir S52915 S529 | nitrate reductase alpha chain - Bacillus subtilis (fragment)   | 79    | 58      | 156         |
| 2940      | 1      | 768        | 385       | gi 149429       | putative [Lactococcus lactis]  | 79    | 72      | 384         |
| 2946      | 1      | 570        | 286       | gi 143267       | 2-oxoglutarate dehydrogenase (odhA; EC 1.2.4.2) [Bacillus subtilis]  | 79    | 61      | 285         |
| 2999      | 1      | 3          | 212       | gi 710020       | nitrite reductase (nirB) [Bacillus subtilis]   | 79    | 59      | 210         |
| 3022      | 1      | 514        | 332       | gi 450686       | 3-phosphoglycerate kinase [Thermotoga maritima]  | 79    | 61      | 183         |
| 3064      | 1      | 3          | 314       | gi 1204436      | pyruvate formate-lyase [Haemophilus influenzae]  | 79    | 60      | 312         |
| 3083      | 1      | 2          | 220       | gi 1149662      | hlypD gene product [Clostridium perfringens]   | 79    | 56      | 219         |
| 3126      | 1      | 701        | 411       | gi 133950       | large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]   | 79    | 55      | 291         |
| 3181      | 1      | 607        | 326       | gi 133950       | large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]   | 79    | 59      | 282         |
| 3345      | 1      | 3          | 476       | gi 871784       | Cip-like ATP-dependent protease binding subunit [Bos taurus]   | 79    | 63      | 474         |
| 3718      | 1      | 536        | 270       | pir C36889 C368 | leuB protein, inactive - Lactococcus lactis subsp. lactis (strain IL1403)  | 79    | 71      | 267         |
| 3724      | 2      | 159        | 401       | gi 1009366      | Respiratory nitrate reductase [Bacillus subtilis]  | 79    | 64      | 243         |
| 3836      | 1      | 608        | 312       | gi 1524193      | unknown [Mycobacterium tuberculosis]   | 79    | 65      | 297         |
| 3941      | 1      | 2          | 334       | gi 415855       | deoxyribose aldolase [Mycoplasma hominis]  | 79    | 54      | 333         |
| 4117      | 1      | 3          | 341       | gi 143015       | gluconate kinase [Bacillus subtilis]   | 79    | 63      | 339         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 4501      | 1      | 406        | 209       | gi1022726       | unknown [Staphylococcus haemolyticus]   | 79    | 66      | 198         |
| 4612      | 1      | 2          | 238       | gi1460689       | TVC [Thermotomomyces vulgaris]  | 79    | 58      | 237         |
| 2         | 1      | 2          | 1213      | gi1520753       | DNA topoisomerase I [Bacillus subtilis]   | 78    | 64      | 1212        |
| 8         | 2      | 2266       | 1220      | gi1216151       | DNA polymerase (gene L; tlg start codon) [Bacteriophage SP021 gi1579197 SP02 DNA polymerase (aa 1-648) [Bacteriophage SP021 p1r[A21499]DJBP52 DNA-directed DNA polymerase (EC 2.7.7.7) - phage P02] | 78    | 72      | 1047        |
| 9         | 2      | 1340       | 1089      | gi1064787       | function unknown [Bacillus subtilis]  | 78    | 57      | 252         |
| 32        | 6      | 6803       | 7702      | gi146974        | NHJ-dependent NAD synthetase [Escherichia coli]   | 78    | 63      | 900         |
| 36        | 4      | 2941       | 3138      | gi1290503       | glutamate permease [Escherichia coli]   | 78    | 53      | 198         |
| 53        | 15     | 17684      | 16221     | gi1303941       | VqIV [Bacillus subtilis]  | 78    | 58      | 1464        |
| 57        | 14     | 10520      | 12067     | gi1072418       | glcA gene product [Staphylococcus carnosus]   | 78    | 65      | 1548        |
| 66        | 7      | 6798       | 5812      | gi1212229       | VqBV [Bacillus subtilis]  | 78    | 67      | 987         |
| 67        | 4      | 4029       | 4376      | gi1466612       | nlkA [Escherichia coli]   | 78    | 71      | 348         |
| 91        | 9      | 10058      | 10942     | gi1467380       | stage 0 sporulation [Bacillus subtilis]   | 78    | 50      | 885         |
| 102       | 12     | 8574       | 10130     | gi149426        | putative [Lactococcus lactis]   | 78    | 61      | 1557        |
| 112       | 6      | 3540       | 4463      | gi1854234       | cymC gene product [Klebsiella oxytoca]  | 78    | 56      | 924         |
| 124       | 2      | 1888       | 1061      | gi1405622       | unknown [Bacillus subtilis]   | 78    | 60      | 828         |
| 130       | 3      | 1805       | 2260      | gi1256636       | putative [Bacillus subtilis]  | 78    | 71      | 456         |
| 133       | 1      | 751        | 377       | gi168060        | lamB [Escherichia nidulans]   | 78    | 59      | 375         |
| 166       | 4      | 7125       | 6163      | gi1451216       | mannosephosphate isomerase [Streptococcus mutans]   | 78    | 63      | 963         |
| 186       | 1      | 1586       | 795       | gi1289284       | cytateinyl-tRNA synthetase [Bacillus subtilis]  | 78    | 63      | 792         |
| 195       | 4      | 2749       | 2315      | gi1353874       | unknown [Rhodobacter capsulatus]  | 78    | 58      | 435         |
| 199       | 3      | 4279       | 3623      | gi143525        | succinate dehydrogenase cytochrome b-558 subunit [Bacillus subtilis] p1r[A29843]DESSC succinate dehydrogenase (EC 1.3.99.1) cytochrome 558 - Bacillus subtilis                                      | 78    | 57      | 657         |
| 199       | 4      | 7209       | 5557      | gi1462521       | deoxyribodipyrimidine photolyase [Bacillus subtilis] p1r[A27192]A27192 uvrB protein - Bacillus subtilis sp14951JUVRC_BACSU EXCINUCLEASE ABC SUBUNIT C   | 78    | 62      | 1653        |
| 223       | 3      | 3831       | 3523      | gi139596        | [Escherichia coli 15200 insertion sequence from ECOR63, partial].l. ene product [Escherichia coli]  | 78    | 47      | 309         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 299       | 4      | 1865       | 2149      | gi1467439       | temperature sensitive cell division [Bacillus subtilis]  | 78    | 62      | 285         |
| 321       | 9      | 7734       | 7735      | gi1142979       | ORF3 is homologous to an ORF downstream of the spot gene of E.coli; RF3 [Bacillus stearothermophilus]  | 78    | 55      | 420         |
| 352       | 4      | 3714       | 3944      | gi1349050       | actin 1 (Pneumocystis carinii)   | 78    | 42      | 231         |
| 352       | 5      | 7592       | 8093      | gi1903587       | NADH dehydrogenase subunit 5 [Bacillus subtilis] sp[P39755]NDHF_BACSU NADH DEHYDROGENASE SUBUNIT 5 (EC 1.6.5.3) NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5 | 78    | 58      | 1500        |
| 376       | 1      | 2          | 583       | gi1551693       | dethiobiotin synthase [Bacillus sphaericus]  | 78    | 34      | 582         |
| 424       | 2      | 1595       | 1768      | gi11524117      | alpha-acetolactate decarboxylase [Lactococcus lactis]  | 78    | 68      | 174         |
| 450       | 1      | 1914       | 1988      | gi1030068       | NAD(P)H oxidoreductase, isoflavone reductase homologue [Solanum tuberosum]   | 78    | 63      | 927         |
| 558       | 1      | 762        | 562       | gi11511588      | bi-functional protein [Methanococcus jannaschii]   | 78    | 60      | 201         |
| 670       | 3      | 1152       | 1589      | gi11122759      | unknown [Bacillus subtilis]  | 78    | 64      | 438         |
| 714       | 1      | 64         | 732       | gi1143460       | 137 kd minor sigma factor (rhoF, sigB; ttg start codon) [Bacillus subtilis]  | 78    | 57      | 669         |
| 814       | 1      | 3          | 368       | gi11377833      | unknown [Bacillus subtilis]  | 78    | 59      | 366         |
| 981       | 1      | 1381       | 692       | gi1143802       | Gerc2 [Bacillus subtilis]  | 78    | 64      | 690         |
| 995       | 2      | 978        | 727       | gi1296947       | luridine kinase [Escherichia coli]   | 78    | 64      | 252         |
| 1045      | 1      | 3          | 401       | gi11407784      | orf-1; novel antigen [Staphylococcus aureus]   | 78    | 61      | 399         |
| 1163      | 2      | 368        | 186       | gi1410117       | glutaminolactate decarboxylase [Bacillus subtilis]   | 78    | 54      | 183         |
| 2191      | 1      | 794        | 399       | gi1215098       | exclusionase [Bacteriophage 154a]  | 78    | 65      | 396         |
| 2933      | 1      | 2          | 181       | gi11204436      | pyruvate formate-lyase [Haemophilus influenzae]  | 78    | 73      | 180         |
| 3041      | 2      | 129        | 317       | gi1624632       | ClcL [Escherichia coli]  | 78    | 53      | 189         |
| 3581      | 1      | 105        | 401       | gi1763186       | [3-ketoacyl-coA thiolase [Saccharomyces cerevisiae]  | 78    | 55      | 297         |
| 3709      | 1      | 3          | 230       | gi1460689       | TVG [Thermotoga maritima]  | 78    | 58      | 228         |
| 3974      | 1      | 528        | 365       | gi1558839       | unknown [Bacillus subtilis]  | 78    | 65      | 264         |
| 3980      | 1      | 3          | 401       | gi139956        | ITGlc [Bacillus subtilis]  | 78    | 62      | 399         |
| 4056      | 1      | 647        | 354       | gi11256635      | [dihydroxy-acid dehydratase [Bacillus subtilis]  | 78    | 55      | 294         |
| 4114      | 1      | 630        | 316       | gi1509372 5093  | hypothetical protein - Trypanosoma brucei  | 78    | 62      | 315         |
| 4185      | 1      | 3          | 179       | gi11339950      | large subunit of NADH-dependent glutamate synthase [Plectononema boryanum]   | 78    | 58      | 177         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 4235      | 1      | 655        | 129       | gi 558839       | unknown [Bacillus subtilis]   | 78    | 60      | 327         |
| 4352      | 1      | 541        | 102       | gi 603768       | Hucl protein, imidazole-5-propionate hydrolase [Bacillus subtilis]<br>gi 603768 Hucl protein, imidazole-5-propionate hydrolase [Bacillus subtilis]          | 78    | 63      | 240         |
| 4388      | 1      | 612        | 307       | gi 1353678      | heavy-metal transporting P-type ATPase [Proteus mirabilis]  | 78    | 59      | 306         |
| 4461      | 1      | 428        | 216       | gi 1276841      | glutamate synthase (COGAT) [Porphyra purpurea]  | 78    | 36      | 213         |
| 4530      | 1      | 474        | 238       | gi 39956        | iglc [Bacillus subtilis]  | 78    | 65      | 237         |
| 3         | 2      | 2969       | 2073      | gi 1109684      | ProV [Bacillus subtilis]  | 77    | 56      | 897         |
| 12        | 2      | 2426       | 1965      | gi 467335       | ribosomal protein L9 [Bacillus subtilis]  | 77    | 59      | 462         |
| 27        | 1      | 2          | 388       | gi 1212728      | YqhI [Bacillus subtilis]  | 77    | 63      | 387         |
| 39        | 2      | 590        | 1252      | gi 40054        | phenylalanyl-tRNA synthetase beta subunit (AA 1-804) [Bacillus subtilis]  | 77    | 60      | 663         |
| 42        | 6      | 2704       | 2931      | gi 606241       | 30S ribosomal subunit protein S14 [Escherichia coli] sp P02370 RS14_ECOLI<br>30S RIBOSOMAL PROTEIN S14, [SUB 2-101]   | 77    | 65      | 228         |
| 46        | 18     | 15459      | 16622     | gi 297798       | mitochondrial formate dehydrogenase precursor [Solanum tuberosum]<br>pir JQ272 JQ272 formate dehydrogenase (EC 1.2.1.2) precursor,<br>ltochondrial - potato | 77    | 55      | 1164        |
| 100       | 4      | 4562       | 4002      | gi 1340128      | ORF1 [Staphylococcus aureus]  | 77    | 54      | 561         |
| 102       | 8      | 5378       | 5713      | gi 1311482      | acetolactate synthase [Thermus aquaticus]   | 77    | 57      | 336         |
| 109       | 7      | 4742       | 5383      | gi 710637       | Unknown [Bacillus subtilis]   | 77    | 56      | 642         |
| 117       | 1      | 2          | 1228      | gi 1337015      | ORF4 [Bacillus subtilis]  | 77    | 53      | 1227        |
| 124       | 10     | 8323       | 7688      | gi 405819       | thymidine kinase [Bacillus subtilis]  | 77    | 63      | 636         |
| 147       | 3      | 1146       | 985       | gi 849027       | hypothetical 15.9-kDa protein [Bacillus subtilis]   | 77    | 37      | 162         |
| 152       | 10     | 7354       | 7953      | gi 1205383      | spermidine/putrescine transport ATP-binding protein [Haemophilus influenzae]  | 77    | 55      | 600         |
| 149       | 2      | 1004       | 1282      | gi 473825       | elongation factor EF-2a' [Escherichia coli]   | 77    | 58      | 279         |
| 184       | 2      | 380        | 1147      | gi 216314       | esterase [Bacillus stearothermophilus]  | 77    | 60      | 768         |
| 189       | 7      | 3296       | 3668      | gi 853809       | ORF3 [Clostridium perfringens]  | 77    | 48      | 573         |
| 193       | 1      | 132        | 290       | gi 1303788      | Yqeh [Bacillus subtilis]  | 77    | 54      | 159         |
| 195       | 8      | 8740       | 8414      | gi 1495820      | M. jannaschii predicted coding region M0798 [Methanococcus jannaschii]  | 77    | 44      | 327         |
| 205       | 8      | 5428       | 5204      | gi 216340       | ORF for adenylate kinase [Bacillus subtilis]  | 77    | 61      | 225         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 205       | 29     | 14795      | 14502     | gi 786155       | Ribosomal Protein L23 [Bacillus subtilis]   | 77    | 62      | 294         |
| 211       | 5      | 1908       | 2084      | gi 410132       | OMP8 [Bacillus subtilis]  | 77    | 47      | 177         |
| 217       | 5      | 3478       | 4416      | gi 496254       | fibronectin/fibrinogen-binding protein [Streptococcus pyogenes]   | 77    | 54      | 939         |
| 232       | 1      | 267        | 998       | gi 1407784      | orf-1; novel antigen [Staphylococcus aureus]  | 77    | 57      | 732         |
| 233       | 2      | 1819       | 1346      | gi 467408       | unknown [Bacillus subtilis]   | 77    | 61      | 474         |
| 243       | 3      | 2661       | 2299      | gi 516155       | unconventional myosin [Sus scrofa]  | 77    | 32      | 363         |
| 299       | 1      | 68         | 769       | gi 467436       | unknown [Bacillus subtilis]   | 77    | 54      | 702         |
| 301       | 4      | 1468       | 1283      | gi 950071       | ATP-bind. pyrimidine kinase [Mycoplasma capricolum] pir S48605 S48605 hypothetical protein - Mycoplasma capricolum SCC3 (fragment)  | 77    | 48      | 186         |
| 302       | 5      | 2741       | 3211      | gi 508980       | phsB [Bacillus subtilis]  | 77    | 57      | 471         |
| 302       | 7      | 3835       | 4863      | gi 147783       | euvs protein [Escherichia coli]   | 77    | 60      | 1029        |
| 307       | 9      | 5402       | 4797      | gi 1070015      | protein-dependent [Bacillus subtilis]   | 77    | 60      | 606         |
| 312       | 1      | 99         | 1391      | gi 143165       | malic enzyme (EC 1.1.1.38) [Bacillus stearothermophilus] pir A33307 DEBSXS malate dehydrogenase oxaloacetate-decarboxylating (EC 1.1.1.38) - Bacillus stearothermophilus    | 77    | 62      | 1293        |
| 312       | 2      | 1541       | 2443      | gi 139855       | carboxyltransferase beta subunit [Synechococcus PCC7942]  | 77    | 58      | 903         |
| 321       | 5      | 5666       | 4596      | gi 39844        | fumarate (citG) [aa 1-462] [Bacillus subtilis]  | 77    | 65      | 1071        |
| 354       | 1      | 47         | 568       | gi 154634       | YmaB [Bacillus subtilis]  | 77    | 57      | 522         |
| 365       | 1      | 2          | 1021      | gi 143374       | phosphoribosyl glycineamide synthetase (PUR-D; gtc start codon) [Bacillus subtilis]   | 77    | 62      | 1020        |
| 374       | 1      | 1          | 708       | gi 1405446      | transketolase [Bacillus subtilis]   | 77    | 61      | 708         |
| 385       | 1      | 1128       | 565       | gi 533099       | endonuclease III [Bacillus subtilis]  | 77    | 63      | 564         |
| 392       | 2      | 594        | 1940      | gi 556014       | UDP-N-acetyl muramate-alanine ligase [Bacillus subtilis] sp P40778 HUKC_UDP-N-ACETYLURAMATE--ALANINE LIGASE (EC 3.2.8) (UDP-N-ACETYLURAMATE--ALANINE SYNTHETASE) (FRAGMENT) | 77    | 65      | 1347        |
| 405       | 5      | 4079       | 3570      | gi 1303912      | Yqmw [Bacillus subtilis]  | 77    | 64      | 510         |
| 487       | 4      | 1302       | 1472      | gi 432427       | ORF1 gene product [Acinetobacter calcoaceticus]   | 77    | 48      | 171         |
| 522       | 1      | 2          | 562       | pir A01179 SYNS | tyrosine-tRNA ligase (EC 6.1.1.1) - Bacillus stearothermophilus   | 77    | 63      | 561         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match Accession | Match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 523       | 2      | 1587       | 1351      | gi11387979      | 44% identity over 302 residues with hypothetical protein from <i>Synechocystis</i> sp. accession D64006.CD; expression induced by environmental stress; some similarity to glycyl transferase; two potential membrane-spanning helices [Bacillus subtilis] | 77    | 48      | 237         |
| 536       | 2      | 983        | 612       | gi1143366       | adenylosuccinate lyase (PUB-B) [Bacillus subtilis] pir[C29326]MB505 adenylosuccinate lyase (EC 4.3.2.2) - Bacillus subtilis  | 77    | 61      | 372         |
| 548       | 2      | 339        | 872       | gi1143387       | aspartate transcarbamylase [Bacillus subtilis]   | 77    | 56      | 534         |
| 597       | 1      | 2          | 483       | gi1904198       | hypothetical protein [Bacillus subtilis]   | 77    | 33      | 480         |
| 633       | 2      | 1747       | 1313      | gi1387577       | ORF1A [Bacillus subtilis]  | 77    | 64      | 435         |
| 642       | 1      | 45         | 360       | gi146971        | euip gene product [Staphylococcus epidermidis]   | 77    | 61      | 276         |
| 659       | 1      | 125        | 1219      | gi11072181      | glutamyl-aminopeptidase [Lactococcus lactis]   | 77    | 62      | 1095        |
| 670       | 4      | 1587       | 1820      | gi1122760       | unknown [Bacillus subtilis]  | 77    | 58      | 234         |
| 789       | 1      | 2          | 391       | gi1137823       | aminopeptidase [Bacillus subtilis]   | 77    | 65      | 390         |
| 815       | 1      | 10         | 573       | gi11303861      | YggW [Bacillus subtilis]   | 77    | 49      | 564         |
| 849       | 1      | 1          | 225       | gi11201844      | H. influenzae predicted coding region H10594 [Haemophilus influenzae]  | 77    | 55      | 225         |
| 1083      | 1      | 3          | 188       | gi1460828       | B969 [Saccharomyces cerevisiae]  | 77    | 66      | 186         |
| 1942      | 1      | 415        | 209       | gi1160047       | p101/acidic basic repeat antigen (Plasmodium falciparum) p101/A29332/A29332 101K malaria antigen precursor - Plasmodium falciparum (strain Camp)   | 77    | 38      | 207         |
| 2559      | 1      | 1          | 171       | gi11499034      | M. jamaensis predicted coding region M0255 [Methanococcus jamaensis]   | 77    | 61      | 171         |
| 2933      | 2      | 243        | 401       | gi142370        | pyruvate formate-lyase (AA 1-760) [Escherichia coli] ir[S01788]S01788 formate C-acetyltransferase (EC 2.3.1.54) - Escherichia coli   | 77    | 72      | 159         |
| 2966      | 1      | 56         | 292       | gi11524397      | glycine betaine transporter OpuD [Bacillus subtilis]   | 77    | 45      | 237         |
| 2976      | 1      | 614        | 309       | gi140003        | oxoglutarate dehydrogenase (NADP+) [Bacillus subtilis] p123129/OD01_BACSU 2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT (EC 2.4.2) (ALPHA-KETOGLUTARATE DEHYDROGENASE)   | 77    | 60      | 306         |
| 2979      | 2      | 678        | 400       | gi11204354      | spore germination and vegetative growth protein [Haemophilus influenzae]   | 77    | 61      | 279         |
| 2988      | 1      | 601        | 377       | gi1438465       | Probable operon with orf. Possible alternative initiation codon, aaas 2151-2153. Homology with acetyltransferases; putative Bacillus subtilis  | 77    | 55      | 225         |
| 2990      | 1      | 331        | 167       | gi1142562       | ATP synthase epsilon subunit [Bacillus megaterium] pir[B28599]MB5EM H-transporting ATP synthase (EC 3.6.1.34) epsilon chain - Bacillus megaterium  | 77    | 63      | 165         |
| 3032      | 1      | 3          | 389       | gi1468430       | alcohol dehydrogenase 2 [Entamoeba histolytica]  | 77    | 56      | 387         |
| 3057      | 1      | 1          | 195       | gi1468764       | mock gene product [Rhizobium meliloti]   | 77    | 50      | 195         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 4008      | 1      | 726        | 400       | gi 603768       | Hut1 protein, imidazolone-5-propionate hydrolase [Bacillus subtilis]<br>gi 603768 Hut1 protein, imidazolone-5-propionate hydrolase [Bacillus subtilis]  | 77    | 52      | 327         |
| 4048      | 1      | 703        | 386       | gi 216278       | [Gramicidin S synthetase 1 [Bacillus brevis]  | 77    | 55      | 318         |
| 4110      | 1      | 3          | 368       | gi 552915 S29   | nitrate reductase alpha chain - Bacillus subtilis (fragment)  | 77    | 61      | 366         |
| 4115      | 1      | 1          | 348       | gi 517205       | 167 kDa Myosin-crossreactive streptococcal antigen [Streptococcus yogenes]  | 77    | 65      | 348         |
| 4225      | 1      | 590        | 297       | gi 1322245      | mevalonate pyrophosphate decarboxylase [Mettus norvegicus]  | 77    | 60      | 294         |
| 4611      | 2      | 494        | 327       | gi 508979       | GTP-binding protein [Bacillus subtilis]   | 77    | 57      | 168         |
| 466A      | 1      | 361        | 182       | gi 552915 S29   | nitrate reductase alpha chain - Bacillus subtilis (fragment)  | 77    | 61      | 180         |
| 25        | 1      | 2          | 1627      | gi 1150620      | HnsA [Streptococcus pneumoniae]   | 76    | 58      | 1626        |
| 3A        | 5      | 1488       | 2537      | gi A43577 A435  | regulatory protein pfor - Clostridium perfringens   | 76    | 57      | 1050        |
| 52        | 5      | 2962       | 4041      | gi 1161061      | dioxigenase [Methylobacterium extorquens]   | 76    | 62      | 1080        |
| 56        | 20     | 27389      | 27955     | gi 467402       | unknown [Bacillus subtilis]   | 76    | 56      | 567         |
| 57        | 15     | 12046      | 12219     | gi 1206040      | weak similarity to keratin [Caenorhabditis elegans]   | 76    | 40      | 174         |
| 91        | 2      | 1062       | 2261      | gi 475715       | acetyl coenzyme A acetyltransferase (thiolase) [Clostridium acetobutylicum]   | 76    | 57      | 1200        |
| 9A        | 2      | 818        | 1624      | gi 467422       | unknown [Bacillus subtilis]   | 76    | 62      | 807         |
| 9A        | 5      | 2965       | 3228      | gi 469793       | y98 gene product [pediococcus acidilactici]   | 76    | 52      | 264         |
| 9A        | 8      | 5922       | 6326      | gi 467427       | methionyl-tRNA synthetase [Bacillus subtilis]   | 76    | 53      | 405         |
| 104       | 3      | 1322       | 1885      | gi 1216151      | DNA polymerase (gene L; ttg start codon) [Bacteriophage SP02] gi 579197<br>SP02 DNA polymerase (aa 1-648) [Bacteriophage SP02] pir A21498 DUBPS2 DNA-<br>directed DNA polymerase [EC 2.7.7.7] - phage P02 | 76    | 63      | 564         |
| 124       | 9      | 8134       | 7055      | gi 853776       | peptide chain release factor 1 [Bacillus subtilis] pir S55437 S55437<br>peptide chain release factor 1 - Bacillus subtilis  | 76    | 58      | 1080        |
| 160       | 5      | 2832       | 3311      | gi 1204976      | prolyl-tRNA synthetase [Haemophilus influenzae]   | 76    | 53      | 480         |
| 160       | 2      | 2617       | 1841      | gi 1177253      | putative ATP-binding protein of ABC-type [Bacillus subtilis]  | 76    | 58      | 777         |
| 189       | 2      | 163        | 888       | gi 467384       | unknown [Bacillus subtilis]   | 76    | 63      | 726         |
| 235       | 3      | 2253       | 3518      | gi 142936       | [polyl-polyglutamate synthetase [Bacillus subtilis] pir B40646 B40646 folC -<br>Bacillus subtilis]  | 76    | 53      | 1266        |
| 236       | 1      | 335        | 925       | gi 1146197      | putative [Bacillus subtilis]  | 76    | 54      | 591         |
| 237       | 8      | 5323       | 5541      | gi 1279261      | [F3G3.6 [Caenorhabditis elegans]  | 76    | 47      | 219         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name   | % sim | % ident | Length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 263       | 5      | 5490       | 4585      | gi 1510348      | [dihydrodipicolinate synthase (Methanococcus jannaschii)]   | 76    | 49      | 906         |
| 304       | 3      | 1051       | 1794      | gi 666982       | [putative membrane spanning subunit (Bacillus subtilis) pir[S52382/S52382] probable membrane spanning protein - Bacillus subtilis]    | 76    | 60      | 744         |
| 312       | 4      | 3611       | 4624      | gi 143312       | [6-phospho-1-fructokinase (gtg start codon; EC 2.7.1.11) (Bacillus tharophilus)]  | 76    | 56      | 1014        |
| 343       | 1      | 2          | 1036      | gi 405956       | [yaeE (Escherichia coli)]   | 76    | 59      | 1035        |
| 347       | 1      | 409        | 1701      | gi 396304       | [acetylornithine deacetylase (Escherichia coli)]  | 76    | 72      | 1293        |
| 358       | 1      | 672        | 1907      | gi 1146215      | [39.0% identity to the Escherichia coli S1 ribosomal protein; putative (Bacillus subtilis)]   | 76    | 58      | 1236        |
| 371       | 1      | 1          | 222       | gi 537084       | [alternate gene name mgt; CG Site No. 497 (Escherichia coli) pir[S56468/S56468] mgtA protein - Escherichia coli]                      | 76    | 61      | 222         |
| 379       | 4      | 4331       | 4858      | gi 143288       | [dihydroliponamide transsuccinylase (odh; EC 2.3.1.61) (Bacillus subtilis)]   | 76    | 61      | 528         |
| 404       | 5      | 4022       | 4492      | gi 3302823      | [yqfC (Bacillus subtilis)]  | 76    | 60      | 471         |
| 411       | 1      | 2          | 307       | gi 86025        | [ORF YKL027W (Saccharomyces cerevisiae)]  | 76    | 55      | 306         |
| 472       | 3      | 4356       | 2854      | gi 1405464      | [Alat (Bacillus subtilis)]  | 76    | 57      | 1503        |
| 546       | 1      | 273        | 995       | gi 153821       | [streptococcal pyrogenic exotoxin type C (speci precursor Streptococcus pyogenes)]  | 76    | 36      | 723         |
| 588       | 1      | 1054       | 557       | gi 1002520      | [MutS (Bacillus subtilis)]  | 76    | 61      | 498         |
| 591       | 1      | 16         | 735       | gi 885934       | [CipB (Synechococcus sp.)]  | 76    | 44      | 720         |
| 602       | 2      | 175        | 798       | gi 1486422      | [OppD homologue (Rhizobium sp.)]  | 76    | 52      | 624         |
| 619       | 2      | 547        | 290       | gi 330613       | [major capsid protein (Human cytomegalovirus)]  | 76    | 47      | 258         |
| 660       | 4      | 2568       | 3302      | gi 904199       | [hypothetical protein (Bacillus subtilis)]  | 76    | 55      | 735         |
| 677       | 1      | 452        | 228       | gi 40177        | [spoof gene product (Bacillus subtilis)]  | 76    | 58      | 225         |
| 962       | 1      | 24         | 206       | gi 143443       | [adenylosuccinate synthetase (Bacillus subtilis) sp[P29726] PURA BACSU ADENYLOSUCCINATE SYNTHETASE (EC 6.3.4.4) IMP-ASPARTATE LIGASE] | 76    | 67      | 183         |
| 978       | 1      | 1158       | 580       | gi 1511333      | [M. jannaschii predicted coding region M3322 (Methanococcus jannaschii)]  | 76    | 56      | 579         |
| 997       | 1      | 486        | 244       | gi 467154       | [No definition line found (Mycobacterium leprae)]   | 76    | 38      | 243         |
| 1563      | 1      | 529        | 266       | gi 1303984      | [yqg (Bacillus subtilis)]   | 76    | 52      | 264         |
| 2184      | 1      | 361        | 182       | gi 506706       | [CapD (Staphylococcus aureus)]  | 76    | 38      | 180         |
| 2572      | 1      | 1          | 387       | gi 153898       | [transport protein (Salmonella typhimurium)]  | 76    | 65      | 387         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 2942      | 1      | 29         | 400       | gi 710020       | nitrite reductase (nirB) (Bacillus subtilis)  | 76    | 59      | 372         |
| 2957      | 1      | 377        | 216       | gi 1511251      | hypothetical protein ISP-P2404 (Methanococcus jannaschii)   | 76    | 47      | 162         |
| 2980      | 1      | 554        | 279       | gi 1405464      | Alar (Bacillus subtilis)  | 76    | 53      | 276         |
| 3015      | 1      | 649        | 326       | gi 408115       | ornithine acetyltransferase (Bacillus subtilis)   | 76    | 61      | 324         |
| 3124      | 1      | 13         | 174       | gi 882705       | ORF_0401 (Escherichia coli)   | 76    | 65      | 162         |
| 3179      | 1      | 3          | 161       | gi 168477       | ferredoxin-dependent glutamate synthase (zoo mayes) pfr[AJ8596]AJ8596 glutamate synthase (ferredoxin) (EC 1.4.7.1) - alze | 76    | 53      | 159         |
| 3789      | 1      | 2          | 379       | gi 39956        | IIIGlc (Bacillus subtilis)  | 76    | 55      | 378         |
| 3892      | 1      | 3          | 314       | gi 1510398      | ferrityochelin binding protein (Methanococcus jannaschii)   | 76    | 52      | 312         |
| 3928      | 1      | 798        | 400       | gi 143016       | permease (Bacillus subtilis)  | 76    | 59      | 399         |
| 4159      | 1      | 757        | 386       | sp P80544 MRSP_ | METHICILLIN-RESISTANT SURFACE PROTEIN (FRAGMENTS)   | 76    | 66      | 372         |
| 4204      | 1      | 17         | 331       | gi 296464       | ATPase (Lactococcus lactis)   | 76    | 56      | 315         |
| 4394      | 1      | 494        | 249       | gi 987255       | Menkes disease gene (Homo sapiens)  | 76    | 48      | 246         |
| 4506      | 1      | 2          | 313       | gi 216746       | D-lactate dehydrogenase (Lactobacillus plantarum)   | 76    | 47      | 312         |
| 4546      | 1      | 477        | 247       | gi 1339950      | large subunit of NADH-dependent glutamate synthase (Plectonema boryanum)  | 76    | 61      | 231         |
| 4596      | 1      | 379        | 191       | gi 560027       | cellulose synthase (Acetobacter xylinum)  | 76    | 70      | 189         |
| 4         | 5      | 5257       | 4337      | gi 882532       | ORF_0294 (Escherichia coli)   | 75    | 59      | 921         |
| 6         | 1      | 164        | 952       | gi 40960        | ORCase (Escherichia coli)   | 75    | 56      | 789         |
| 12        | 3      | 5935       | 3944      | gi 467336       | unknown (Bacillus subtilis)   | 75    | 57      | 1992        |
| 23        | 18     | 18272      | 17310     | gi 2296433      | O-acetylserine sulphydrase B (Alcaligenes eutrophus)  | 75    | 55      | 963         |
| 25        | 3      | 2356       | 3393      | gi 1502419      | Plex (Bacillus subtilis)  | 75    | 56      | 1038        |
| 36        | 8      | 5765       | 6037      | gi 2256317      | unknown (Schizosaccharomyces pombe)   | 75    | 45      | 273         |
| 46        | 13     | 11186      | 12058     | gi 48972        | nitrate transporter (Synchococcus sp.)  | 75    | 46      | 873         |
| 51        | 7      | 3474       | 3677      | gi 113607       | sporulation protein (Bacillus subtilis)   | 75    | 61      | 204         |
| 53        | 16     | 16850      | 16590     | gi 143402       | recombination protein (ttg start codon) (Bacillus subtilis) gi 1303923 RecW (Bacillus subtilis)                           | 75    | 51      | 261         |
| 74        | 3      | 3572       | 2568      | gi 1204847      | ornithine carbamoyltransferase (Haemophilus influenzae)   | 75    | 61      | 1005        |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 85        | 3      | 4628       | 3930      | gi1143368       | [phosphoribosylformyl glycylamide synthetase I (PUR-1); gcg start odon] (Bacillus subtilis)           | 75    | 63      | 699         |
| 85        | 5      | 5588       | 4878      | gi1143367       | [phosphoribosyl aminodazole succinocarboxamide synthetase (PUR-C; tg start odon)] (Bacillus subtilis) | 75    | 55      | 711         |
| 85        | 8      | 6625       | 7530      | gi11303916      | YqIA (Bacillus subtilis)  | 75    | 53      | 906         |
| 87        | 3      | 2340       | 3590      | gi11064813      | [homologous to sp:PHOR_BACSU (Bacillus subtilis)]   | 75    | 56      | 1251        |
| 87        | 6      | 6084       | 6896      | gi11064810      | [function unknown (Bacillus subtilis)]  | 75    | 61      | 813         |
| 108       | 2      | 1844       | 1503      | gi11001824      | [hypothetical protein (Synchocystis sp.)  | 75    | 51      | 342         |
| 110       | 3      | 1748       | 3727      | gi11447593      | [putative pPOp synthetase (Streptomyces coelicolor)]  | 75    | 55      | 1980        |
| 110       | 7      | 4151       | 5252      | gi11177251      | [cldD gene product (Bacillus subtilis)]   | 75    | 75      | 900         |
| 120       | 34     | 11266      | 110649    | gi11526394      | [ORF-2 upstream of gbaB operon (Bacillus subtilis)]   | 75    | 55      | 618         |
| 121       | 5      | 2050       | 4221      | gi11154632      | [NrdE (Bacillus subtilis)]  | 75    | 54      | 2172        |
| 124       | 3      | 283        | 143       | gi1405622       | [unknown (Bacillus subtilis)]   | 75    | 56      | 141         |
| 128       | 1      | 81         | 1139      | gi1143316       | [genI gene products (Bacillus megaterium)]  | 75    | 48      | 1059        |
| 130       | 8      | 5760       | 5903      | gi11256654      | [54.RV identity with Neisseria gonorrhoeae regulatory protein piliD; putative (Bacillus subtilis)]    | 75    | 62      | 143         |
| 136       | 2      | 4480       | 3185      | gi1467403       | [seryl-tRNA synthetase (Bacillus subtilis)]   | 75    | 54      | 1296        |
| 161       | 10     | 5439       | 5798      | gi11001195      | [hypothetical protein (Synchocystis sp.)  | 75    | 55      | 360         |
| 172       | 4      | 3819       | 2995      | gi1735153       | [ATP-binding protein (Bacillus subtilis)]   | 75    | 52      | 825         |
| 179       | 1      | 2024       | 1107      | gi1143037       | [porphobilinogen deaminase (Bacillus subtilis)]   | 75    | 58      | 918         |
| 195       | 10     | 9529       | 9374      | gi10257451      | [HYPOTHETICAL PROTEIN IN PURB 5' REGION (ORF-15) (FRAGMENT)]  | 75    | 60      | 156         |
| 200       | 4      | 2805       | 4596      | gi1142440       | [ATP-dependent nuclease (Bacillus subtilis)]  | 75    | 56      | 1992        |
| 206       | 3      | 6900       | 5620      | gi11256135      | YbbF (Bacillus subtilis)  | 75    | 53      | 1281        |
| 216       | 2      | 159        | 389       | gi11052800      | [unknown (Schizosaccharomyces pombe)]   | 75    | 58      | 231         |
| 229       | 1      | 29         | 847       | gi11205958      | [branched chain aa transport system II carrier protein (Haemophilus influenzae)]                      | 75    | 49      | 819         |
| 230       | 2      | 518        | 1714      | gi1971337       | [nitrite extrusion protein (Bacillus subtilis)]   | 75    | 53      | 1197        |
| 231       | 1      | 2240       | 1122      | gi11002521      | [MucB (Bacillus subtilis)]  | 75    | 56      | 1119        |
| 233       | 3      | 1314       | 1859      | gi1467405       | [unknown (Bacillus subtilis)]   | 75    | 59      | 546         |

TABLE 2

TABLE 2 - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession        | Match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|------------------------|--|-------|---------|-------------|
| 269       | 1      | 325        | 164       | [gi1511246]            | [methyl coenzyme M reductase system, component A2 [Methanococcus jannaschii]]  | 75    | 50      | 162         |
| 292       | 1      | 1389       | 772       | [gi1511604]            | [M. jannaschii predicted coding region MJ1651 [Methanococcus jannaschii]]  | 75    | 46      | 618         |
| 304       | 4      | 1773       | 2281      | [gi1205328]            | [surfacein [Haemophilus influenzae]]   | 75    | 55      | 489         |
| 312       | 3      | 2437       | 3387      | [gi1285621]            | [undefined open reading frame [Bacillus stearothermophilus]]   | 75    | 62      | 951         |
| 312       | 5      | 4622       | 6403      | [gi1041097]            | [Pyruvate Kinase [Bacillus psychrophilus]]   | 75    | 57      | 1782        |
| 319       | 1      | 353        | 877       | [gi1212728]            | [YqhI [Bacillus subtilis]]   | 75    | 54      | 525         |
| 320       | 5      | 4321       | 5031      | [gi1070361]            | [OMP decarboxylase [Lactococcus lactis]]   | 75    | 56      | 711         |
| 320       | 6      | 5010       | 5642      | [gi1143394]            | [OMP-PRPP transferase [Bacillus subtilis]]   | 75    | 60      | 633         |
| 337       | 4      | 1519       | 2088      | [gi1487433]            | [citrate synthase II [Bacillus subtilis]]  | 75    | 58      | 570         |
| 394       | 2      | 669        | 1271      | [gi1304976]            | [matches P50017: ATP_GTP_A and P500301: EFAC108_GTP: similar to longation factor G. TetM/TetO tetracycline-resistance proteins Escherichia coli]   | 75    | 51      | 603         |
| 423       | 1      | 127        | 570       | [gi11183839]           | [unknown [Pseudomonas aeruginosa]]   | 75    | 59      | 444         |
| 433       | 2      | 1603       | 1929      | [gi1149211]            | [acetolactate synthase [Klebsiella pneumoniae]]  | 75    | 63      | 327         |
| 446       | 2      | 176        | 1540      | [gi1312441]            | [dihydroorotase [Bacillus caldolyticus]]   | 75    | 62      | 1365        |
| 486       | 1      | 494        | 249       | [gi1149682]            | [potF gene product [Clostridium perfringens]]  | 75    | 55      | 246         |
| 496       | 1      | 3          | 794       | [gi1143582]            | [spoIIIEA protein [Bacillus subtilis]]   | 75    | 59      | 792         |
| 498       | 2      | 824        | 1504      | [gi1143328]            | [phoP protein (put.): putative [Bacillus subtilis]]  | 75    | 47      | 681         |
| 499       | 2      | 1061       | 1624      | [gi1138799]            | [44% identity over 302 residues with hypothetical protein from Synchocystis sp. accession D6406.C0; expression induced by environmental stress: some similarity to glycosyl transferases: two potential membrane-spanning helices [Bacillus subtilis]] | 75    | 51      | 564         |
| 568       | 1      | 641        | 653       | [ptr134110]JC41 (SCC3) | [triacylglycerol lipase (EC 3.1.1.3) 2 - Mycoplasma mycoides subsp. mycoides (SCC3)]   | 75    | 50      | 189         |
| 613       | 2      | 430        | 233       | [gi1320993]            | [tegument protein [Salivarius herpavirus 2]]   | 75    | 75      | 198         |
| 621       | 1      | 1          | 525       | [gi1529754]            | [speC [Streptococcus pyogenes]]  | 75    | 43      | 523         |
| 642       | 5      | 1809       | 2474      | [gi1176401]            | [EpiG [Staphylococcus epidermidis]]  | 75    | 51      | 666         |
| 646       | 2      | 454        | 657       | [gi1172442]            | [ribonuclease P [Saccharomyces cerevisiae]]  | 75    | 37      | 204         |
| 657       | 1      | 3          | 147       | [gi1862541]            | [OMP_0236 [Escherichia coli]]  | 75    | 47      | 345         |
| 750       | 1      | 1662       | 832       | [gi146971]             | [epiP gene product [Staphylococcus epidermidis]]   | 75    | 57      | 831         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 754       | 1      | 2          | 481       | gi 1303901      | YnfP [Bacillus subtilis]   | 75    | 57      | 480         |
| 763       | 2      | 563        | 393       | gi 1205145      | multidrug resistance protein [Haemophilus influenzae]  | 75    | 51      | 171         |
| 775       | 1      | 951        | 482       | pir B36889 B368 | leuA protein, inactive - Lactococcus lactis subsp. lactis (strain IL1403)  | 75    | 63      | 480         |
| 793       | 1      | 1          | 180       | gi 143316       | [gap] gene products [Bacillus megaterium]  | 75    | 57      | 180         |
| 800       | 1      | 318        | 160       | gi 509411       | NFRA protein [Azorhizobium caulinodans]  | 75    | 34      | 159         |
| 811       | 1      | 1117       | 560       | gi 143634       | Rho Factor [Bacillus subtilis]   | 75    | 60      | 558         |
| 940       | 1      | 493        | 329       | gi 1276985      | arginase [Bacillus caldovelox]   | 75    | 50      | 165         |
| 971       | 2      | 37         | 252       | gi 1001373      | hypothetical protein [Synecocystis sp.]  | 75    | 58      | 216         |
| 1059      | 1      | 384        | 232       | gi 726480       | L-glutamine-D-fructose-6-phosphate amidotransferase [Bacillus subtilis]  | 75    | 67      | 153         |
| 1109      | 2      | 219        | 374       | gi 143331       | alkaline phosphatase regulatory protein phoR - Bacillus subtilis<br>spi P23345 PHOR_BACSU ALKALINE PHOSPHATASE SYNTHESIS SENSOR PROTEIN PHOR (EC 2.7.3.-)  | 75    | 53      | 156         |
| 1268      | 1      | 271        | 137       | gi 304135       | ornithine acetyltransferase [Bacillus stearothermophilus]<br>spi Q07908 ARG3_BACST GLUTAMATE N-ACETYLTRANSFERASE (EC 2.3.1.35) ORNITHINE ACETYLTRANSFERASE [ORNITHINE TRANSACETYLASE] (OATASE) / HIND-ACID ACETYLTRANSFERASE (EC 2.3.1.1) (N-ACETYLGLUTAMATE YNTIA | 75    | 63      | 135         |
| 1500      | 1      | 324        | 163       | gi 1205488      | lexinuclease ABC subunit B [Haemophilus influenzae]  | 75    | 57      | 162         |
| 1529      | 1      | 798        | 400       | gi 1002521      | MutL [Bacillus subtilis]   | 75    | 54      | 399         |
| 3010      | 1      | 770        | 387       | gi 1204435      | pyruvate formate-lyase activating enzyme [Haemophilus influenzae]  | 75    | 54      | 384         |
| 3105      | 1      | 1          | 180       | gi 1041097      | pyruvate kinase [Bacillus psychrophilus]   | 75    | 57      | 180         |
| 3117      | 1      | 45         | 212       | gi 899317       | peptide synthetase module [Microcystis aeruginosa] pir S49111 S49111<br>probable amino acid activating domain - Microcystis aeruginosa (fragment) (SUB 144-528)  | 75    | 42      | 168         |
| 3139      | 2      | 139        | 345       | gi 145294       | adenine phosphoribosyl-transferase [Escherichia coli]  | 75    | 66      | 207         |
| 3880      | 1      | 618        | 310       | gi 1009366      | Respiratory nitrate reductase [Bacillus subtilis]  | 75    | 58      | 309         |
| 3911      | 1      | 48         | 401       | gi 433991       | ATP synthase subunit beta [Bacillus subtilis]  | 75    | 68      | 354         |
| 3957      | 1      | 2          | 379       | pir B36889 B368 | 3-isopropylmalate dehydratase (EC 4.2.1.33) chain leuC - Lactococcus lactis subsp. lactis (strain IL1403)  | 75    | 65      | 378         |
| 4005      | 1      | 5          | 259       | gi 216746       | D-lactate dehydrogenase [Lactobacillus plantarum]  | 75    | 48      | 255         |
| 4080      | 1      | 73         | 333       | gi 415855       | deoxyribose aldolase [Mycoplasma hominis]  | 75    | 59      | 261         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | ncycp gene name   | V sim | V ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 4111      | 1      | 1          | 339       | gi1169435       | putative [Lactococcus lactis]   | 75    | 57      | 339         |
| 4136      | 1      | 602        | 303       | gi1450688       | hsdM gene of Ecopri gene product [Escherichia coli] pir[S08437]S08437 hsdM protein - Escherichia coli pir[S05625]S05625 hypothetical protein A - Escherichia coli (S08 40-520)  | 75    | 56      | 300         |
| 4144      | 1      | 668        | 336       | gi148972        | nitrate transporter [Synecococcus sp.]  | 75    | 49      | 333         |
| 4237      | 1      | 664        | 374       | gi1133950       | large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]  | 75    | 55      | 291         |
| 4306      | 2      | 73         | 318       | gi1294260       | major surface glycoprotein [Pneumocystis carinii]   | 75    | 68      | 246         |
| 4343      | 1      | 715        | 359       | gi11204652      | methylated-DNA--protein-cysteine methyltransferase [Haemophilus influenzae]   | 75    | 52      | 357         |
| 4532      | 1      | 620        | 312       | gi1396464       | ATPase [Lactococcus lactis]   | 75    | 55      | 309         |
| 38        | 9      | 5776       | 6126      | gi1443793       | HupC [Escherichia coli]   | 74    | 50      | 351         |
| 50        | 8      | 6910       | 6221      | gi1123988       | hypothetical protein [Bacillus subtilis]  | 74    | 55      | 690         |
| 56        | 9      | 10770      | 12221     | gi11000451      | Trap [Bacillus subtilis]  | 74    | 57      | 1452        |
| 64        | 2      | 2266       | 1622      | gi141015        | anpartate-trna ligase [Escherichia coli]  | 74    | 57      | 645         |
| 66        | 6      | 5063       | 4848      | gi11212729      | YqhJ [Bacillus subtilis]  | 74    | 47      | 216         |
| 67        | 18     | 14334      | 14897     | gi11510631      | endoglucanase [Methanococcus jannaschii]  | 74    | 52      | 564         |
| 102       | 15     | 12561      | 13136     | gi1149429       | putative [Lactococcus lactis]   | 74    | 67      | 576         |
| 102       | 16     | 13121      | 14419     | gi1149435       | putative [Lactococcus lactis]   | 74    | 57      | 1299        |
| 108       | 4      | 4873       | 3902      | gi139478        | ATP binding protein of transport ATPase [Bacillus firmus] iZ1515486[S15486 ATP-binding protein - Bacillus firmus p125946]YATR_BACF1 HYPOTHETICAL ATP-BINDING TRANSPORT PROTEIN. | 74    | 59      | 972         |
| 116       | 5      | 8574       | 7093      | gi11205430      | dipeptide transport system permease protein [Haemophilus influenzae]  | 74    | 49      | 1482        |
| 120       | 7      | 4342       | 4803      | gi1146970       | ribonucleoside triphosphate reductase [Escherichia coli] pir[A4733]A4733 anaerobic ribonucleotide reductase - Escherichia coli  | 74    | 58      | 462         |
| 121       | 7      | 5961       | 6581      | gi11107528      | ctg start [Campylobacter coli]  | 74    | 51      | 621         |
| 128       | 3      | 2320       | 3531      | gi1143318       | phosphoglycerate kinase [Bacillus megaterium]   | 74    | 57      | 1212        |
| 130       | 7      | 5237       | 5791      | gi11256653      | DNA-binding protein [Bacillus subtilis]   | 74    | 60      | 555         |
| 136       | 3      | 6745       | 5150      | gi1143076       | histidase [Bacillus subtilis]   | 74    | 58      | 1596        |
| 145       | 2      | 664        | 1368      | gi1407773       | devA gene product [Anabaena sp.]  | 74    | 45      | 705         |
| 152       | 1      | 552        | 277       | gi11377833      | unknown [Bacillus subtilis]   | 74    | 54      | 276         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 164       | 10     | 11064      | 11375     | gi1580900       | ORF3 gene product [Bacillus subtilis]  | 74    | 52      | 312         |
| 175       | 2      | 3109       | 2624      | gi1642636       | unknown [Rhizobium meliloti]   | 74    | 34      | 486         |
| 175       | 9      | 6064       | 5612      | gi1854656       | [Na/H] antiporter system ORF2 [Bacillus alcalophilus]  | 74    | 46      | 453         |
| 195       | 11     | 111346     | 10339     | gi1204430       | hypothetical protein (SP:P23745) [Haemophilus influenzae]  | 74    | 55      | 1008        |
| 205       | 17     | 9619       | 9059      | gi11044979      | ribosomal protein L6 [Bacillus subtilis]   | 74    | 64      | 561         |
| 216       | 7      | 5574       | 6710      | gi11146207      | putative [Bacillus subtilis]   | 74    | 63      | 1137        |
| 241       | 3      | 4521       | 3134      | gi1694121       | malate thiokinase [Methylobacterium extorquens]  | 74    | 52      | 1188        |
| 246       | 6      | 3305       | 2799      | gi1467374       | single strand DNA binding protein [Bacillus subtilis]  | 74    | 64      | 507         |
| 249       | 4      | 6551       | 5313      | gi11524397      | glycine betaine transporter OpuD [Bacillus subtilis]   | 74    | 55      | 1319        |
| 261       | 7      | 4389       | 4081      | gi1809542       | ChrB protein [Erwinia chrysanthemi]  | 74    | 42      | 309         |
| 278       | 6      | 5714       | 4665      | gi1204872       | ATP-binding protein [Haemophilus influenzae]   | 74    | 54      | 1050        |
| 309       | 1      | 1220       | 666       | gi1205579       | hypothetical protein (GB:U14003.102) [Haemophilus influenzae]  | 74    | 53      | 555         |
| 315       | 2      | 1473       | 862       | gi1143398       | quinol oxidase [Bacillus subtilis]   | 74    | 57      | 612         |
| 320       | 1      | 1          | 1065      | gi1943389       | glutaminase of carbamyl phosphate synthetase [Bacillus subtilis]<br>pir[E19845]E19845 carbamyl-phosphate synthase glutamine-hydrolyzing [EC 6.3.5.5], pyrimidine-repressible, small hain - Bacillus subtilis | 74    | 60      | 1065        |
| 380       | 2      | 382        | 1128      | gi1534857       | ATPase subunit a [Bacillus stearothermophilus]   | 74    | 56      | 747         |
| 405       | 2      | 1742       | 1311      | gi1303915       | YqhZ [Bacillus subtilis]   | 74    | 65      | 432         |
| 433       | 5      | 2503       | 3270      | gi1473502       | alpha-acetolactate synthase [Lactococcus lactis]   | 74    | 56      | 768         |
| 452       | 1      | 1          | 942       | gi1413982       | lpa-58r gene product [Bacillus subtilis]   | 74    | 52      | 942         |
| 461       | 1      | 3          | 1193      | gi1558494       | homoserine dehydrogenase [Bacillus subtilis]   | 74    | 51      | 1191        |
| 461       | 2      | 1174       | 1407      | gi140211        | threonine synthase (thrC) (AA 1-352) [Bacillus subtilis] ir[A25364]<br>threonine synthase (EC 4.2.99.2) - Bacillus btilla  | 74    | 56      | 234         |
| 462       | 2      | 402        | 736       | gi1142520       | thioredoxin [Bacillus subtilis]  | 74    | 62      | 333         |
| 478       | 1      | 574        | 320       | gi11499005      | glycyl-tRNA synthetase [Methanococcus jannaschii]  | 74    | 52      | 255         |
| 501       | 2      | 719        | 1740      | gi1217040       | acid glycoprotein [Streptococcus pyogenes]   | 74    | 58      | 1002        |
| 551       | 2      | 4083       | 2791      | gi143040        | glutamate-1-semialdehyde 2,1-aminotransferase [Bacillus subtilis]<br>pir[D12728]D12728 glutamate-1-semialdehyde 2,1-aminotransferase (EC 4.3.8) - Bacillus subtilis  | 74    | 51      | 1293        |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 573       | 1      | 1          | 477       | gi1006605       | hypothetical protein [Synecocystis sp.]   | 74    | 45      | 477         |
| 596       | 2      | 1780       | 1298      | gi11303853      | YqgP [Bacillus subtilis]  | 74    | 55      | 483         |
| 618       | 2      | 2924       | 1758      | gi11146237      | 21.4% of identity to trans-acting transcription factor of <i>Saccharomyces cerevisiae</i> ; 25% of identity to sucrose synthase of <i>Zea mays</i> ; putative [Bacillus subtilis] | 74    | 55      | 1187        |
| 659       | 2      | 1269       | 1595      | gi11072380      | ORF3 [Lactococcus lactis]   | 74    | 62      | 327         |
| 724       | 1      | 373        | 188       | gi1143374       | [phosphoribosyl] glycinnamide synthetase (PUR-D, gtp start codon) <i>Bacillus subtilis</i>  | 74    | 58      | 186         |
| 743       | 2      | 604        | 1209      | gi1153813       | ORF1, putative [Streptococcus parasanguis]  | 74    | 50      | 606         |
| 836       | 1      | 2          | 259       | gi1143458       | ORF V [Bacillus subtilis]   | 74    | 47      | 258         |
| 989       | 2      | 443        | 724       | gi11503994      | YqkM [Bacillus subtilis]  | 74    | 46      | 282         |
| 1106      | 1      | 1          | 492       | gi1146970       | lepid gene product [Staphylococcus epidermidis]   | 74    | 54      | 492         |
| 1135      | 2      | 373        | 528       | gi1143348       | lpa-24d gene product [Bacillus subtilis]  | 74    | 48      | 156         |
| 1234      | 1      | 817        | 452       | gi11495245      | lrecJ gene product [Erwinia chrysanthemi]   | 74    | 36      | 366         |
| 2586      | 1      | 2          | 238       | gi11149701      | lsbcC gene product [Clostridium perfringens]  | 74    | 62      | 237         |
| 2959      | 1      | 798        | 400       | gi11405454      | aconitase [Bacillus subtilis]   | 74    | 60      | 399         |
| 2962      | 1      | 650        | 363       | gi11450686      | 3-phosphoglycerate kinase [Thermotoga maritima]   | 74    | 58      | 288         |
| 2983      | 1      | 1          | 191       | gi11303893      | YqgH [Bacillus subtilis]  | 74    | 56      | 189         |
| 3018      | 1      | 2          | 223       | gi1143040       | glutamate-1-semialdehyde 2,1-aminotransferase [Bacillus subtilis]<br>pir10427281042728 glutamate-1-semialdehyde 2,1-aminotransferase (EC 4.3.8) -<br>[Bacillus subtilis]          | 74    | 56      | 222         |
| 3038      | 1      | 510        | 256       | pir1529151529   | nitrate reductase alpha chain - <i>Bacillus subtilis</i> (fragment)   | 74    | 57      | 255         |
| 3062      | 1      | 374        | 189       | gi11107528      | lctg start [Campylobacter coli]   | 74    | 51      | 186         |
| 4035      | 1      | 184        | 360       | gi11022725      | unknown [Staphylococcus haemolyticus]   | 74    | 64      | 177         |
| 4045      | 1      | 607        | 305       | gi11510977      | M. jannaschii predicted coding region M0938 [Methanococcus jannaschii]  | 74    | 41      | 303         |
| 4283      | 1      | 471        | 304       | gi1520844       | lorf4 [Bacillus subtilis]   | 74    | 58      | 168         |
| 4449      | 1      | 3          | 221       | gi1580910       | peptide-synthetase ORF1 [Bacillus subtilis]   | 74    | 54      | 219         |
| 4587      | 1      | 458        | 231       | gi11370207      | lorf6 [Lactobacillus sake]  | 74    | 59      | 228         |

TABLE 2

5. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 4603      | 1      | 23         | 214       | gi1146208       | glutamate synthase large subunit (EC 2.6.1.53) [Escherichia coli] pir[A29617]A29617 glutamate synthase (NADPH) (EC 1.4.1.13) large hain - Escherichia coli | 74    | 60      | 186         |
| 4670      | 1      | 366        | 184       | gi1256135       | ybbP [Bacillus subtilis]   | 74    | 61      | 183         |
| 5         | 10     | 7953       | 7162      | gi1343727       | putative [Bacillus subtilis]   | 73    | 42      | 792         |
| 11        | 2      | 2454       | 1372      | gi1366338       | dihydroxotrate dehydrogenase [Agrobacterium ageritae]  | 73    | 55      | 1083        |
| 14        | 1      | 2024       | 1020      | gi1343373       | [phosphoribosyl aminoimidazole carboxy formyl transferase/inosine monophosphate cyclohydrolase (Pur-His)] Bacillus subtilis                                | 73    | 54      | 1005        |
| 23        | 5      | 5426       | 4635      | gi13468339      | meso-2,3-butanediol dehydrogenase (D-acetoin forming) [Klebsiella pneumoniae]  | 73    | 58      | 792         |
| 23        | 17     | 17379      | 16360     | gi1297060       | ornithine cyclodeaminase [Rhizobium meliloti]  | 73    | 37      | 1020        |
| 29        | 2      | 692        | 1273      | gi1467442       | [stage V sporulation [Bacillus subtilis]   | 73    | 54      | 582         |
| 31        | 5      | 6467       | 4914      | gi1414000       | [ipa-76d gene product [Bacillus subtilis]  | 73    | 55      | 1554        |
| 37        | 8      | 8658       | 7402      | gi1429259       | [pepT gene product [Bacillus subtilis]   | 73    | 59      | 1257        |
| 37        | 9      | 7738       | 7562      | gi1468367       | [alpha-isopropylmalate isomerase (put.) putative [Rhizomucor ircinelloides]  | 73    | 52      | 177         |
| 38        | 7      | 3931       | 4896      | gi1405885       | [yeiN [Escherichia coli]   | 73    | 58      | 966         |
| 44        | 6      | 5041       | 4238      | gi1540895       | [unknown [Bacillus subtilis]   | 73    | 53      | 804         |
| 44        | 11     | 7767       | 8306      | gi142009        | [moaB gene product [Escherichia coli]  | 73    | 50      | 540         |
| 45        | 3      | 2439       | 3080      | gi1109685       | [proW [Bacillus subtilis]  | 73    | 47      | 642         |
| 54        | 13     | 14016      | 13794     | gi1413931       | [ipa-7d gene product [Bacillus subtilis]   | 73    | 61      | 243         |
| 59        | 4      | 1430       | 2248      | gi1147923       | [threonine dehydratase 2 (EC 4.2.1.16) [Escherichia coli]  | 73    | 53      | 819         |
| 65        | 1      | 1458       | 730       | gi1677944       | [appF [Bacillus subtilis]  | 73    | 56      | 729         |
| 80        | 2      | 1375       | 860       | gi1580932       | [murD gene product [Bacillus subtilis]   | 73    | 53      | 516         |
| 102       | 13     | 10124      | 11179     | gi1580891       | [3-isopropylmalate dehydrogenase (AA 1 - 365) [Bacillus subtilis] pir[A26522]A26522 3-isopropylmalate dehydrogenase (EC 1.1.1.85) - acillus subtilis       | 73    | 55      | 1056        |
| 109       | 2      | 3493       | 2600      | gi1510849       | [M. jannaschii predicted coding region M0775 [Methanococcus jannaschii]  | 73    | 40      | 894         |
| 120       | 8      | 4782       | 5756      | gi146970        | [ribonucleoside triphosphate reductase [Escherichia coli] pir[A47331]A47331 anaerobic ribonucleotide reductase - Escherichia coli                          | 73    | 56      | 975         |
| 120       | 9      | 5726       | 6223      | gi1204333       | [anaerobic ribonucleoside-triphosphate reductase [Haemophilus influenzae]  | 73    | 62      | 498         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 132       | 5      | 4151       | 4363      | gi 871048       | [HPSK2 - heavy chain potential motor protein (Giardia intestinalis)]   | 73    | 43      | 213         |
| 140       | 6      | 5952       | 4324      | gi 634107       | [kdpB (Escherichia coli)]  | 73    | 59      | 1629        |
| 142       | 6      | 7040       | 5919      | gi 410125       | [rfrI gene product (Bacillus subtilis)]  | 73    | 57      | 1122        |
| 149       | 4      | 1866       | 1717      | gi 460892       | [heparin binding protein-44, HBP-44 (mice, peptide, 360 aa)<br>pir JX0281 JX0281 heparin-binding protein-44 precursor - mouse gi 220434<br>ORF (Mus musculus) (SUB 2-360)] | 73    | 53      | 150         |
| 158       | 1      | 1          | 1431      | gi 882504       | [ORF_f560 (Escherichia coli)]  | 73    | 57      | 1431        |
| 174       | 6      | 5352       | 4525      | gi 1146240      | [ketopantoate hydroxymethyltransferase (Bacillus subtilis)]  | 73    | 55      | 828         |
| 175       | 8      | 5537       | 5178      | gi 854657       | [Na/H antiporter system ORF3 (Bacillus alcalophilus)]  | 73    | 56      | 360         |
| 186       | 5      | 6593       | 5493      | gi 467477       | [unknown (Bacillus subtilis)]  | 73    | 48      | 1101        |
| 249       | 6      | 6283       | 5729      | gi 1524397      | [glycine betaine transporter OpuB (Bacillus subtilis)]   | 73    | 56      | 555         |
| 265       | 4      | 1873       | 2280      | gi 39848        | [j03 (Bacillus subtilis)]  | 73    | 41      | 408         |
| 270       | 1      | 328        | 582       | gi 780461       | [320 kDa polypeptide (African swine fever virus)]  | 73    | 53      | 255         |
| 278       | 4      | 4283       | 3618      | gi 1204965      | [hypothetical 23.3 kD protein (Escherichia coli)]  | 73    | 49      | 666         |
| 279       | 3      | 4984       | 3593      | gi 1385288      | [isochorismate synthase (Bacillus subtilis)]   | 73    | 58      | 1392        |
| 291       | 4      | 1207       | 1575      | gi 1511440      | [glutamine--fructose-6-phosphate transaminase (Methanococcus jannaschii)]  | 73    | 63      | 369         |
| 299       | 2      | 735        | 1166      | gi 467437       | [unknown (Bacillus subtilis)]  | 73    | 58      | 432         |
| 299       | 5      | 2050       | 3234      | gi 467439       | [temperature sensitive cell division (Bacillus subtilis)]  | 73    | 53      | 1185        |
| 314       | 1      | 1237       | 728       | gi 536655       | [ORF_YBR244W (Saccharomyces cerevisiae)]   | 73    | 43      | 510         |
| 336       | 2      | 1827       | 1036      | gi 790943       | [urea amidolyase (Bacillus subtilis)]  | 73    | 51      | 792         |
| 374       | 3      | 1389       | 1874      | gi 1405451      | [ynaJ (Bacillus subtilis)]   | 73    | 55      | 486         |
| 433       | 4      | 1916       | 2554      | gi 473902       | [alpha-acetolactate synthase (Lactococcus lactis)]   | 73    | 54      | 639         |
| 509       | 2      | 1795       | 1028      | gi 467483       | [unknown (Bacillus subtilis)]  | 73    | 56      | 768         |
| 513       | 1      | 1709       | 918       | gi 1146220      | [NAD+ dependent glycerol-3-phosphate dehydrogenase (Bacillus subtilis)]  | 73    | 56      | 792         |
| 533       | 2      | 239        | 733       | gi 1510605      | [hypothetical protein (SP1942297) (Methanococcus jannaschii)]  | 73    | 44      | 495         |
| 546       | 2      | 1148       | 2815      | gi 41748        | [hscM protein (AA 1-520) (Escherichia coli)]   | 73    | 52      | 1668        |
| 549       | 1      | 762        | 382       | gi 1314847      | [ClnA (Bacillus subtilis)]   | 73    | 57      | 381         |
| 567       | 1      | 1346       | 675       | gi 410137       | [ORFX13 (Bacillus subtilis)]   | 73    | 58      | 672         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 716       | 2      | 654        | 1112      | gi 1256623      | exodeoxyribonuclease (Bacillus subtilis)  | 73    | 56      | 459         |
| 772       | 1      | 3          | 677       | gi 142010       | Shows 70.3% similarity and 48.6% identity to the ENVH protein of almonella typhimurium (Anabaena sp.)   | 73    | 57      | 675         |
| 774       | 1      | 3          | 209       | gi 409286       | bmrU (Bacillus subtilis)  | 73    | 52      | 207         |
| 782       | 1      | 1          | 402       | gi 143320       | [gap] gene products (Bacillus megaterium)   | 73    | 56      | 402         |
| 789       | 2      | 451        | 762       | gi 1063246      | low homology to p14 protein of Haemophilus influenzae and 14.2 kDa protein of Escherichia coli (Bacillus subtilis)  | 73    | 56      | 312         |
| 796       | 1      | 3          | 911       | gi 453754       | ABC transporter (Bacillus subtilis)   | 73    | 58      | 908         |
| 806       | 3      | 1209       | 949       | gi 143786       | cryptophenyl-tRNA synthetase (EC 6.1.1.2) (Bacillus subtilis) pir J70481 TMS tryptophan-tRNA ligase (EC 6.1.1.2) - Bacillus ubtills                               | 73    | 51      | 261         |
| 816       | 2      | 4839       | 3097      | gi 41748        | hdm protein (AA 1-520) (Escherichia coli)   | 73    | 52      | 1743        |
| 839       | 1      | 798        | 400       | gi 886906       | argininosuccinate synthetase (Streptomyces clavuligerus) pir S37659 S37659 argininosuccinate synthase (EC 6.3.4.5) - Streptomyces clavuligerus                    | 73    | 59      | 399         |
| 857       | 1      | 3          | 290       | gi 348052       | acetoin utilization protein (Bacillus subtilis)   | 73    | 50      | 288         |
| 1008      | 1      | 790        | 398       | gi 40100        | rodC (tag3) polypeptide (AA 1-746) (Bacillus subtilis) Ir S06049 S06049 rodC protein - Bacillus subtilis p P13485 TAGF_BACSU TECHOIC ACID BIOSYNTHESIS PROTEIN F. | 73    | 41      | 393         |
| 1018      | 1      | 1          | 213       | gi 529357       | No definition line found (Caenorhabditis elegans) sp P46975 STT3_CAEEL OULGOSACCHARYL TRANSFERASE STT3 SUBUNIT OHLOG.   | 73    | 53      | 213         |
| 1031      | 1      | 3          | 491       | gi 142706       | comC gene product (Bacillus subtilis)   | 73    | 51      | 499         |
| 1174      | 1      | 395        | 204       | gi 1149513      | alpha subunit of laminin 5 (Homo sapiens)   | 73    | 60      | 192         |
| 1175      | 1      | 655        | 329       | gi 473817       | "ORF" (Escherichia coli)  | 73    | 57      | 327         |
| 1187      | 1      | 3          | 209       | gi 580870       | lpa-37d qoxA gene product (Bacillus subtilis)   | 73    | 52      | 207         |
| 1206      | 1      | 72         | 245       | gi 144816       | formyltetrahydrofolate synthetase (Trifus) (tfg start codon) (EC 3.4.3) (Moorella thermocetica)   | 73    | 43      | 174         |
| 1454      | 1      | 423        | 241       | gi 113253       | unknown (Schizosaccharomyces pombe)   | 73    | 53      | 183         |
| 1469      | 1      | 517        | 260       | gi 1303787      | YqeG (Bacillus subtilis)  | 73    | 55      | 258         |
| 1761      | 1      | 174        | 189       | gi 9135         | Mst26aa gene product (Drosophila simulans)  | 73    | 34      | 186         |
| 1849      | 1      | 467        | 243       | gi 162307       | DNA topoisomerase II (Trypanosoma cruzi)  | 73    | 60      | 225         |
| 2055      | 1      | 2          | 400       | gi 559281       | P47K protein (Rhodococcus erythropolis)   | 73    | 34      | 399         |
| 2556      | 1      | 2          | 244       | gi 145925       | feca (Escherichia coli)   | 73    | 62      | 243         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 2947      | 2      | 549        | 400       | gi1186680       | polynucleotide phosphorylase [Bacillus subtilis]  | 73    | 51      | 150         |
| 2956      | 1      | 746        | 375       | gi143397        | quinol oxidase [Bacillus subtilis]  | 73    | 58      | 372         |
| 3037      | 1      | 655        | 329       | gi111091        | acetolactate synthase [Bacillus subtilis]   | 73    | 55      | 327         |
| 3115      | 1      | 385        | 194       | gi1323866       | overlapping out-of-phase protein [Egplant mosaic virus]   | 73    | 53      | 192         |
| 3603      | 2      | 700        | 527       | gi11419521      | glutaryl-CoA dehydrogenase precursor [Mus musculus]   | 73    | 48      | 174         |
| 3743      | 1      | 798        | 400       | gi1450688       | hcdh gene of Ecoor1 gene product [Escherichia coli] p1r[S18437/S18437 hcdh protein - Escherichia coli p1r[S09629/S09629 hypothetical protein A - Escherichia coli (SUB 40-520)] | 73    | 54      | 399         |
| 3752      | 1      | 640        | 359       | gi1524193       | unknown [Mycobacterium tuberculosis]  | 73    | 59      | 282         |
| 3852      | 1      | 2          | 181       | gi1216746       | D-lactate dehydrogenase [Lactobacillus plantarum]   | 73    | 68      | 180         |
| 3914      | 1      | 475        | 239       | gi151490/S134   | hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5) - Chicken (fragment)  | 73    | 53      | 237         |
| 3914      | 2      | 570        | 343       | gi1528991       | unknown [Bacillus subtilis]   | 73    | 38      | 228         |
| 4069      | 1      | 2          | 316       | gi140003        | oxoglutarate dehydrogenase (HAPD) [Bacillus subtilis] p1r[S13125/S001_BACSU 2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT (EC 2.4.2) (ALPHA-KETOGLUTARATE DEHYDROGENASE)]           | 73    | 55      | 315         |
| 4165      | 1      | 715        | 365       | gi11439521      | glutaryl-CoA dehydrogenase precursor [Mus musculus]   | 73    | 48      | 351         |
| 4176      | 1      | 1          | 377       | gi1809660       | deoxyribose-phosphate aldolase [Bacillus subtilis] p1r[S49455/S49455 deoxyribose-phosphate aldolase (EC 4.1.2.4) - scillus subtilis]  | 73    | 60      | 177         |
| 4202      | 1      | 572        | 378       | gi1528991       | unknown [Bacillus subtilis]   | 73    | 38      | 195         |
| 4314      | 1      | 2          | 193       | gi1436797       | N-acyl-L-amino acid amidohydrolase [Bacillus stearothermophilus] ep1p17112/ANA_BAGST N-ACYL-L-AMINO ACID AMIDOHYDROLASE (EC 5.1.14) (AMINOACYLASE)]                             | 73    | 47      | 192         |
| 4393      | 1      | 3          | 263       | gi1216267       | ORF2 [Bacillus megaterium]  | 73    | 47      | 261         |
| 35        | 2      | 903        | 1973      | gi11146196      | phosphoglycerate dehydrogenase [Bacillus subtilis]  | 72    | 53      | 1071        |
| 38        | 22     | 19094      | 17877     | gi1602031       | similar to trimethylamine DH [Mycoplasma capricolum] p1r[S49950/S49950 probable trimethylamine dehydrogenase (EC 5.99.7) - Mycoplasma capricolum (SGC3) (fragment)]             | 72    | 54      | 1218        |
| 38        | 23     | 18134      | 19162     | gi1413968       | ipa-44d gene product [Bacillus subtilis]  | 72    | 54      | 1029        |
| 44        | 19     | 11895      | 12953     | gi1516272       | unknown [Bacillus subtilis]   | 72    | 49      | 1059        |
| 48        | 7      | 6248       | 7117      | gi143499        | pyruvate synthase [Halobacterium halobium]  | 72    | 49      | 870         |
| 50        | 7      | 6563       | 5691      | gi11205399      | proton glutamate symport protein [Haemophilus influenzae]   | 72    | 53      | 873         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 53        | 9      | 10521      | 9259      | gi11303956      | YqjE [Bacillus subtilis]   | 72    | 52      | 1263        |
| 56        | 23     | 29549      | 29995     | gi1467471       | unknown [Bacillus subtilis]  | 72    | 47      | 447         |
| 69        | 4      | 5298       | 4123      | gi11354775      | pfos/R [Treponema pallidum]  | 72    | 46      | 1176        |
| 69        | 5      | 4377       | 4982      | gi1904198       | hypothetical protein [Bacillus subtilis]   | 72    | 43      | 606         |
| 73        | 1      | 2          | 856       | gi1142997       | glycerol uptake facilitator [Bacillus subtilis]  | 72    | 59      | 855         |
| 98        | 13     | 9371       | 10258     | gi1467435       | unknown [Bacillus subtilis]  | 72    | 50      | 886         |
| 127       | 1      | 1          | 1593      | gi1217144       | alanine carrier protein [thermophilic bacterium PSJ] pfr[A45111][A45111]   | 72    | 56      | 1593        |
| 131       | 1      | 5197       | 2600      | gi1153952       | polymerase III polymerase subunit (dnaE) [Salmonella typhimurium] pfr[A45915][A45915 DNA-directed DNA polymerase (EC 2.7.7.7)] ttr lpha chain - Salmonella typhimurium | 72    | 53      | 2598        |
| 141       | 4      | 1040       | 1978      | gi11405446      | transketolase [Bacillus subtilis]  | 72    | 54      | 939         |
| 149       | 8      | 2819       | 2535      | gi1606234       | Isacy [Escherichia coli]   | 72    | 44      | 285         |
| 149       | 17     | 5472       | 5245      | gi11304472      | DNA polymerase [unidentified phycodnavirus clone OTU4]   | 72    | 55      | 228         |
| 154       | 1      | 1          | 210       | gi1205620       | ferritin like protein [Haemophilus influenzae]   | 72    | 40      | 210         |
| 155       | 1      | 2207       | 1320      | gi1391610       | [farnesyl diphosphate synthase [Bacillus stearothermophilus] pfr[JX0257][JX0257 geranyltransferase (EC 2.5.1.10)] - Bacillus stearothermophilus                        | 72    | 57      | 884         |
| 180       | 1      | 2          | 328       | gi1431630       | AI80 [Saccharomyces cerevisiae]  | 72    | 62      | 327         |
| 184       | 3      | 1145       | 3553      | gi11205110      | virulence associated protein homolog [Haemophilus influenzae]  | 72    | 49      | 2409        |
| 195       | 2      | 1923       | 1279      | gi11001730      | hypothetical protein [Synecocystis sp.]  | 72    | 45      | 645         |
| 206       | 13     | 14646      | 15869     | gi11064807      | ORTHINE AMINOTRANSFERASE [Bacillus subtilis]   | 72    | 50      | 1224        |
| 209       | 2      | 462        | 932       | gi11204666      | hypothetical protein (GB:X71124_53) [Haemophilus influenzae]   | 72    | 60      | 471         |
| 215       | 2      | 784        | 522       | gi1881513       | insulin receptor homolog [Drosophila melanogaster] pfr[S57245][S57245 insulin receptor homolog - fruit fly (Drosophila melanogaster) (SUB 46-2146)]                    | 72    | 63      | 243         |
| 224       | 1      | 2          | 790       | gi1949974       | sucrose repressor [Staphylococcus xylosum]   | 72    | 54      | 789         |
| 233       | 1      | 1526       | 765       | gi11408493      | homologous to SwissProt:YIDA_ECOLI hypothetical protein [Bacillus subtilis]  | 72    | 52      | 762         |
| 240       | 1      | 220        | 1485      | gi1537049       | ORF_0470 [Escherichia coli]  | 72    | 52      | 1266        |
| 245       | 1      | 3          | 1340      | gi11204578      | hypothetical protein (GB:U06949_1) [Haemophilus influenzae]  | 72    | 46      | 1338        |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 259       | 2      | 2108       | 1245      | gi 1340128      | ORF1 (Staphylococcus aureus)  | 72    | 59      | 864         |
| 304       | 2      | 285        | 1094      | gi 1205130      | glutamine-binding periplasmic protein (Haemophilus influenzae)  | 72    | 52      | 810         |
| 307       | 10     | 3326       | 5039      | gi 1070015      | protein-dependent (Bacillus subtilis)   | 72    | 53      | 286         |
| 315       | 1      | 517        | 260       | gi 143399       | quinol oxidase (Bacillus subtilis)  | 72    | 55      | 258         |
| 316       | 11     | 9672       | 9308      | gi 1204445      | hypothetical protein (SPI27857) (Haemophilus influenzae)  | 72    | 56      | 315         |
| 337       | 3      | 926        | 1609      | gi 487433       | citrate synthase II (Bacillus subtilis)   | 72    | 55      | 684         |
| 364       | 7      | 12538      | 10493     | gi 1510643      | ferrous iron transport protein B (Methanococcus jannaschii)   | 72    | 53      | 2046        |
| 409       | 2      | 340        | 1263      | gi 1402944      | orfR1 gene product (Bacillus subtilis)  | 72    | 49      | 924         |
| 441       | 3      | 2177       | 1590      | gi 312379       | highly conserved among eubacteria (Clostridium acetobutylicum)  | 72    | 48      | 584         |
| 453       | 6      | 2654       | 2505      | gi 500601       | antibacterial protein 3 - Staphylococcus haemolyticus   | 72    | 70      | 150         |
| 460       | 1      | 2          | 625       | gi 1016162      | ABC transporter subunit (Cyanophora paradoxa)   | 72    | 51      | 624         |
| 463       | 1      | 3253       | 1628      | gi 666014       | The polymorphism (RFP) of this gene is associated with susceptibility to essential hypertension. The SA gene product has light homology to acetyl-CoA synthetase (Homo sapiens) | 72    | 60      | 1626        |
| 480       | 4      | 3047       | 3466      | gi 433992       | ATP synthase subunit epsilon (Bacillus subtilis)  | 72    | 53      | 420         |
| 502       | 1      | 1086       | 586       | gi 310859       | ORF2 (Synecococcus sp.)   | 72    | 50      | 501         |
| 519       | 1      | 81         | 1184      | gi 1303704      | YrkE (Bacillus subtilis)  | 72    | 54      | 1104        |
| 559       | 1      | 3          | 746       | gi 1107530      | caud gene product (Campylobacter coli)  | 72    | 56      | 744         |
| 575       | 1      | 1142       | 573       | gi 1303866      | YqoS (Bacillus subtilis)  | 72    | 56      | 570         |
| 671       | 1      | 2          | 592       | gi 1204497      | protein-export membrane protein (Haemophilus influenzae)  | 72    | 44      | 591         |
| 679       | 2      | 295        | 1251      | gi 563258       | virulence-associated protein 2 (Dichelobacter nodosus)  | 72    | 52      | 957         |
| 687       | 2      | 295        | 957       | gi 1166214      | 44% identical amino acids with the Escherichia coli emba suppress; putative (Bacillus subtilis)   | 72    | 49      | 663         |
| 837       | 1      | 1          | 435       | gi 1146183      | putative (Bacillus subtilis)  | 72    | 54      | 435         |
| 868       | 1      | 150        | 788       | gi 1377842      | unknown (Bacillus subtilis)   | 72    | 55      | 639         |
| 922       | 1      | 130        | 432       | gi 1088269      | unknown protein (Atobacter vinelandii)  | 72    | 58      | 303         |
| 941       | 1      | 2          | 238       | gi 153929       | NADPH-sulfite reductase flavoprotein component (Salmonella typhimurium)   | 72    | 49      | 237         |
| 980       | 1      | 840        | 421       | gi 853767       | UDP-N-acetylglucosamine 1-carboxyvinyltransferase (Bacillus subtilis)   | 72    | 59      | 420         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 1209      | 1      | 383        | 213       | gi1144735       | neurotoxin type B [Clostridium botulinum]   | 72    | 44      | 171         |
| 1469      | 2      | 671        | 474       | gi11205458      | hypothetical protein (CB:02562_47) [Haemophilus influenzae]   | 72    | 63      | 198         |
| 1956      | 1      | 727        | 365       | gi1154409       | hexosephosphate transport protein (Salmonella typhimurium) p1r[841853]B41853 hexose phosphate transport system regulatory protein uhpB - Salmonella typhimurium | 72    | 44      | 363         |
| 2101      | 1      | 3          | 401       | gi11301950      | YqjY [Bacillus subtilis]  | 72    | 50      | 399         |
| 2503      | 1      | 569        | 399       | gi1149713       | formate dehydrogenase [Methanobacterium formicicum] p1r[A42712]A42712 formate dehydrogenase [EC 1.2.1.2] - Methanobacterium formicicum                          | 72    | 56      | 171         |
| 2567      | 1      | 3          | 155       | gi11212729      | Yqj3 [Bacillus subtilis]  | 72    | 46      | 153         |
| 3004      | 1      | 367        | 185       | gi11665999      | hypothetical protein [Bacillus subtilis]  | 72    | 55      | 183         |
| 3109      | 1      | 276        | 141       | gi11413968      | ipa-4ed gene product [Bacillus subtilis]  | 72    | 45      | 138         |
| 3171      | 1      | 3          | 287       | gi11519338      | glutamate synthase (ferredoxin) [Synecocystis sp.] p1r[S46957]S46957 glutamate synthase (ferredoxin) [EC 1.4.7.1] - Synecocystis sp.                            | 72    | 52      | 285         |
| 3771      | 1      | 26         | 367       | gi11408501      | homologous to N-acyl-L-amino acid amidohydrolase of Bacillus stearothermophilus [Bacillus subtilis]   | 72    | 63      | 342         |
| 3951      | 1      | 1          | 222       | gi11500409      | M. jannaschii predicted coding region Mj1519 [Methanococcus jannaschii]   | 72    | 38      | 222         |
| 4190      | 1      | 721        | 362       | gi1139956       | 11GIC [Bacillus subtilis]   | 72    | 57      | 360         |
| 4444      | 1      | 3          | 347       | gi11009366      | Respiratory nitrate reductase [Bacillus subtilis]   | 72    | 55      | 345         |
| 6         | 2      | 911        | 1200      | gi11537095      | ornithine carbamoyltransferase [Escherichia coli]   | 71    | 54      | 270         |
| 11        | 15     | 11350      | 10859     | gi11532309      | 35 kDa protein [Escherichia coli]   | 71    | 47      | 492         |
| 19        | 2      | 1248       | 2435      | gi11244574      | D-alanine:D-alanine ligase [Enterococcus hirae]   | 71    | 52      | 1188        |
| 21        | 2      | 898        | 1488      | gi1149629       | anthranilate synthase component 2 [Leptospira biflexa] p1r[C2840]C2840 anthranilate synthase [EC 4.1.3.27] component II Leptospira biflexa                      | 71    | 45      | 591         |
| 34        | 1      | 1          | 567       | gi11301983      | YqjP [Bacillus subtilis]  | 71    | 59      | 567         |
| 37        | 3      | 3192       | 2806      | gi11209681      | glutamate-rich protein [Bacillus firmus]  | 71    | 50      | 387         |
| 38        | 18     | 12250      | 12462     | gi11927645      | arginyl endopeptidase [Porphyromonas gingivalis]  | 71    | 50      | 213         |
| 39        | 3      | 1246       | 4431      | p1r[S09411]S094 | spoIIIE protein - Bacillus subtilis   | 71    | 49      | 3186        |
| 53        | 14     | 15770      | 14760     | gi1142611       | branched chain alpha-keto acid dehydrogenase E1-alpha [Bacillus subtilis]   | 71    | 58      | 1011        |
| 54        | 11     | 13461      | 12625     | gi1143014       | lent repressor [Bacillus subtilis]  | 71    | 46      | 837         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 57        | 7      | 7152       | 5860      | gi 508175       | ETIC domain of PTS-dependent Cat transport and phosphorylation Escherichia coli  | 71    | 48      | 1293        |
| 57        | 18     | 11897      | 14334     | gi 1063247      | high homology to (flavohemoprotein (Haemoglobin-like protein) of Alcaligenes eutrophus and Saccharomyces cerevisiae [Bacillus subtilis])   | 71    | 56      | 438         |
| 62        | 16     | 9831       | 10955     | gi 1303926      | YqjQ [Bacillus subtilis]   | 71    | 54      | 1125        |
| 70        | 12     | 8505       | 8966      | gi 147198       | phnE protein [Escherichia coli]  | 71    | 38      | 462         |
| 86        | 5      | 2394       | 2089      | gi 904205       | hypothetical protein [Bacillus subtilis]   | 71    | 51      | 306         |
| 96        | 7      | 7601       | 8269      | gi 709991       | hypothetical protein [Bacillus subtilis]   | 71    | 49      | 669         |
| 100       | 6      | 4822       | 5931      | gi 1060848      | Opine dehydrogenase [Arthrobacter sp.]   | 71    | 45      | 1110        |
| 103       | 1      | 1062       | 532       | gi 143089       | lep protein [Bacillus subtilis]  | 71    | 41      | 531         |
| 109       | 18     | 15312      | 15695     | gi 413985       | lipa-61d gene product [Bacillus subtilis]  | 71    | 57      | 384         |
| 113       | 1      | 630        | 316       | gi 663254       | probable protein kinase [Saccharomyces cerevisiae]   | 71    | 57      | 315         |
| 114       | 5      | 6598       | 5603      | gi 143156       | membrane bound protein [Bacillus subtilis]   | 71    | 40      | 996         |
| 133       | 2      | 3087       | 1723      | gi 1303913      | YqjX [Bacillus subtilis]   | 71    | 53      | 1365        |
| 149       | 19     | 6335       | 5895      | gi 529650       | G40P [Bacteriophage SP21]  | 71    | 51      | 441         |
| 154       | 5      | 3635       | 3087      | gi 425488       | repressor protein [Streptococcus sobrinus]   | 71    | 47      | 549         |
| 164       | 11     | 11354      | 11689     | gi 149318       | ONF4 gene product [Bacillus subtilis]  | 71    | 52      | 116         |
| 169       | 5      | 1936       | 2745      | gi 1403403      | unknown [Mycobacterium tuberculosis]   | 71    | 56      | 810         |
| 193       | 2      | 272        | 1234      | gi 1303788      | YqjM [Bacillus subtilis]   | 71    | 49      | 963         |
| 205       | 1      | 1743       | 895       | gi 1215694      | GlnQ [Mycoplasma pneumoniae]   | 71    | 46      | 849         |
| 233       | 4      | 1849       | 2022      | gi 633732       | ORF1 [Campylobacter jejuni]  | 71    | 50      | 174         |
| 237       | 7      | 4501       | 5169      | gi 149384       | HisIE [Lactococcus lactis]   | 71    | 54      | 669         |
| 272       | 4      | 2848       | 2273      | gi 709993       | hypothetical protein [Bacillus subtilis]   | 71    | 48      | 576         |
| 274       | 2      | 618        | 1496      | gi 143035       | NAD(P)H:glutamate-transfer RNA reductase [Bacillus subtilis] pif[A35252]A35252 5-aminolevulinate synthase (EC 2.3.1.37) - ecillus subtilis | 71    | 53      | 879         |
| 276       | 5      | 3349       | 2720      | gi 303562       | ORF210 [Escherichia coli]  | 71    | 50      | 630         |
| 287       | 1      | 136        | 660       | gi 310634       | 120 kDa protein [Streptococcus gordonii]   | 71    | 53      | 525         |
| 288       | 6      | 3322       | 2771      | gi 1256625      | putative [Bacillus subtilis]   | 71    | 47      | 552         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 301       | 6      | 3492       | 2461      | gi 467417       | similar to lysine decarboxylase [Bacillus subtilis]  | 71    | 57      | 1032        |
| 306       | 4      | 6607       | 5222      | gi 1256618      | transport protein [Bacillus subtilis]  | 71    | 56      | 1386        |
| 307       | 2      | 1536       | 925       | gi 622683       | orfC [Mycoplasma capricolum]   | 71    | 45      | 612         |
| 310       | 5      | 5793       | 5146      | gi 348052       | acetoin utilization protein [Bacillus subtilis]  | 71    | 51      | 648         |
| 322       | 1      | 2          | 1303      | gi 1001819      | hypothetical protein [Synchocystis sp.]  | 71    | 46      | 1302        |
| 333       | 4      | 4171       | 3995      | gi 167473       | unknown [Bacillus subtilis]  | 71    | 57      | 177         |
| 350       | 2      | 548        | 922       | gi 551879       | ORF 1 [Lactococcus lactis]   | 71    | 55      | 375         |
| 375       | 4      | 1860       | 3071      | gi 467447       | unknown [Bacillus subtilis]  | 71    | 57      | 1212        |
| 380       | 5      | 1560       | 2102      | gi 142557       | ATP synthase b subunit [Bacillus megaterium]   | 71    | 43      | 543         |
| 414       | 2      | 251        | 637       | gi 580904       | homologous to E. coli rnpA [Bacillus subtilis]   | 71    | 49      | 387         |
| 424       | 1      | 335        | 1354      | gi 581305       | l-lactate dehydrogenase [Lactobacillus plantarum]  | 71    | 57      | 1020        |
| 436       | 4      | 3701       | 3270      | pir PM0501 PM05 | phosphoribosylanthranilate isomerase (EC 5.3.1.24) - Bacillus subtilis (fragment)  | 71    | 66      | 432         |
| 482       | 1      | 3          | 1280      | gi 410142       | ORFX18 [Bacillus subtilis]   | 71    | 49      | 1278        |
| 525       | 3      | 2272       | 1844      | gi 143370       | phosphoribosylpyrophosphate amidotransferase (Pur-P; EC 2.4.2.14) Bacillus subtilis  | 71    | 56      | 429         |
| 529       | 4      | 2739       | 2047      | gi 606150       | ORF_1309 [Escherichia coli]  | 71    | 43      | 693         |
| 563       | 1      | 22         | 969       | gi 1237015      | ORF4 [Bacillus subtilis]   | 71    | 53      | 948         |
| 581       | 1      | 506        | 255       | gi 1301730      | 725G3.2 [Caenorhabditis elegans]   | 71    | 47      | 252         |
| 612       | 2      | 1088       | 913       | gi 153968       | fimbriae 2 [Salmonella typhimurium]  | 71    | 55      | 156         |
| 613       | 1      | 1          | 654       | gi 466778       | lysine specific permease [Escherichia coli]  | 71    | 50      | 654         |
| 618       | 1      | 1243       | 623       | gi 1146238      | poly(A) polymerase [Bacillus subtilis]   | 71    | 52      | 623         |
| 630       | 1      | 1170       | 586       | gi 1486243      | unknown [Bacillus subtilis]  | 71    | 53      | 585         |
| 691       | 1      | 1126       | 641       | gi 389260       | comE ORF1 [Bacillus subtilis]  | 71    | 51      | 486         |
| 694       | 2      | 149        | 427       | gi 112971       | NADH dehydrogenase subunit V (AA 1-605) [Callus gallus] ir S10197 S10197 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain - chicken mitochondrion (SOCI) | 71    | 47      | 279         |
| 715       | 2      | 169        | 777       | gi 130830       | YqfL [Bacillus subtilis]   | 71    | 53      | 609         |
| 746       | 2      | 1473       | 970       | gi 1377843      | unknown [Bacillus subtilis]  | 71    | 52      | 504         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 748       | 1      | 1437       | 802       | gi1405459       | YmeS [Bacillus subtilis]   | 71    | 49      | 636         |
| 753       | 1      | 1018       | 524       | gi1510389       | H. Jannaschii predicted coding region MJ0296 [Methanococcus jannaschii]  | 71    | 53      | 495         |
| 761       | 1      | 3          | 215       | gi1475972       | pentafunctional enzyme [Pneumocystis carinii]  | 71    | 47      | 213         |
| 783       | 1      | 1203       | 703       | gi1536655       | ORF YBR244w [Saccharomyces cerevisiae]   | 71    | 52      | 501         |
| 800       | 3      | 1292       | 987       | gi1204326       | tRNA delta(2)-isopentenylpyrophosphate transferase [Haemophilus influenzae]  | 71    | 48      | 306         |
| 806       | 1      | 116        | 286       | gi1419075       | cbiM gene product [Methanobacterium thermoautotrophicum]   | 71    | 50      | 171         |
| 931       | 1      | 973        | 488       | gi1893358       | PgsA [Bacillus subtilis]   | 71    | 56      | 486         |
| 1041      | 1      | 2          | 262       | gi1408507       | pyrimidine nucleoside transport protein [Bacillus subtilis]  | 71    | 45      | 261         |
| 1070      | 1      | 2          | 172       | gi1709993       | hypothetical protein [Bacillus subtilis]   | 71    | 46      | 171         |
| 1176      | 1      | 57         | 365       | gi151359        | HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas mavalonii] pir[A44756/A44756] Pseudomonas sp. hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) | 71    | 49      | 309         |
| 1181      | 1      | 366        | 184       | gi146971        | epiP gene product [Staphylococcus epidermidis]   | 71    | 50      | 183         |
| 1281      | 1      | 3          | 290       | gi153016        | ORF 419 protein [Staphylococcus aureus]  | 71    | 50      | 288         |
| 1348      | 1      | 456        | 229       | gi1402683       | orfC [Mycoplasma capricolum]   | 71    | 48      | 228         |
| 2002      | 1      | 756        | 379       | gi1008177       | ORF YJL016w [Saccharomyces cerevisiae]   | 71    | 48      | 378         |
| 2119      | 1      | 2          | 217       | gi1046088       | arginyl-tRNA synthetase [Mycoplasma genitalium]  | 71    | 50      | 216         |
| 2418      | 1      | 3          | 320       | gi1499771       | H. Jannaschii predicted coding region MJ0936 [Methanococcus jannaschii]  | 71    | 57      | 318         |
| 2961      | 1      | 2          | 187       | gi1312443       | carbamoyl-phosphate synthase (glutamine-hydrolysing) [Bacillus aldolyticus]  | 71    | 57      | 186         |
| 2995      | 2      | 67         | 306       | gi1710020       | nitrite reductase (nirB) [Bacillus subtilis]   | 71    | 43      | 240         |
| 3033      | 1      | 2          | 184       | gi11262335      | YmaA [Bacillus subtilis]   | 71    | 57      | 183         |
| 3584      | 1      | 3          | 338       | gi1401716       | beta-isopropylmalate dehydrogenase [Neurospora crassa]   | 71    | 55      | 336         |
| 3715      | 2      | 743        | 399       | gi1561952       | gluconate permease [Bacillus licheniformis]  | 71    | 59      | 345         |
| 3785      | 1      | 770        | 387       | gi147382        | acyl-CoA-dehydrogenase [Streptomyces purpurascens]   | 71    | 57      | 384         |
| 3875      | 1      | 541        | 272       | gi11001541      | hypothetical protein [Synecocystis sp.]  | 71    | 38      | 270         |
| 4135      | 1      | 637        | 320       | gi1142695       | S-adenosyl-L-methionine:uroporphyrinogen III methyltransferase [Bacillus megaterium]   | 71    | 52      | 318         |
| 4249      | 1      | 63         | 239       | gi11205363      | deoxyribose aldolase [Haemophilus influenzae]  | 71    | 63      | 177         |
| 4508      | 1      | 530        | 267       | gi11197667      | vitellogenin [Anolis pulchellus]   | 71    | 46      | 264         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 6         | 3      | 1237       | 2721      | gi 1321788      | arginine ornithine antiporter [Clostridium perfringens]  | 70    | 54      | 1485        |
| 11        | 11     | 6572       | 7486      | gi 216854       | [P47K] Pseudomonas chlororaphis  | 70    | 41      | 915         |
| 12        | 1      | 2890       | 1481      | gi 467330       | replicative DNA helicase [Bacillus subtilis]   | 70    | 49      | 1410        |
| 15        | 1      | 1756       | 893       | gi 451216       | mannosephosphate isomerase [Streptococcus mutans]  | 70    | 46      | 864         |
| 15        | 2      | 1277       | 1050      | gi 476092       | unknown [Bacillus subtilis]  | 70    | 50      | 228         |
| 17        | 2      | 2132       | 1350      | gi 145402       | choline dehydrogenase [Escherichia coli]   | 70    | 52      | 783         |
| 21        | 1      | 2          | 925       | gi 149516       | anthranilate synthase alpha subunit [Lactococcus lactis] pir S3124/S3124 anthranilate synthase (EC 4.1.3.27) alpha chain - actococcus lactis subsp. lactis | 70    | 50      | 924         |
| 25        | 7      | 5580       | 6251      | gi 1389349      | ORF3 [Bacillus subtilis]   | 70    | 52      | 672         |
| 33        | 6      | 6071       | 7423      | gi 1301875      | YqjB [Bacillus subtilis]   | 70    | 51      | 1353        |
| 36        | 2      | 959        | 1594      | gi 500755       | [methyl] purine glycoylase [Mus musculus]  | 70    | 47      | 636         |
| 38        | 8      | 4901       | 5860      | gi 11408507     | pyrimidine nucleoside transport protein [Bacillus subtilis]  | 70    | 44      | 960         |
| 44        | 8      | 5312       | 5989      | gi 1006620      | hypothetical protein [Synecocystis sp.]  | 70    | 49      | 678         |
| 46        | 10     | 8950       | 10020     | gi 1403126      | csd gene product [Alcaligenes eutrophus]   | 70    | 45      | 1071        |
| 52        | 2      | 2727       | 1900      | gi 1486247      | unknown [Bacillus subtilis]  | 70    | 53      | 828         |
| 52        | 6      | 4048       | 4656      | gi 244501       | esterase II-carboxylesterase (EC 3.1.1.1) [Pseudomonas fluorescens, eptido, 218 aa]  | 70    | 50      | 609         |
| 56        | 8      | 8460       | 9962      | gi 1339951      | small subunit of NADH-dependent glutamate synthase [Plectonema boryanum]   | 70    | 51      | 1503        |
| 62        | 1      | 48         | 290       | gi 142702       | A competence protein 2 [Bacillus subtilis]   | 70    | 47      | 243         |
| 64        | 1      | 1080       | 541       | gi 1204377      | molycoprotein biosynthesis protein [Haemophilus influenzae]  | 70    | 47      | 540         |
| 70        | 5      | 5139       | 3595      | gi 1204834      | [2'-3'-cyclic-nucleotide 2'-phosphodiesterase [Haemophilus influenzae]   | 70    | 47      | 1545        |
| 91        | 4      | 7793       | 5466      | gi 886471       | methionine synthase [Catharanthus roseus]  | 70    | 56      | 2328        |
| 96        | 5      | 8754       | 7255      | pir D9096 D990  | alkaline phosphatase (EC 3.1.3.1) III precursor - Bacillus subtilis  | 70    | 54      | 1500        |
| 110       | 2      | 767        | 1300      | gi 145294       | adenine phosphoribosyl-transferase [Escherichia coli]  | 70    | 51      | 534         |
| 116       | 6      | 7026       | 7976      | gi 143607       | sporulation protein [Bacillus subtilis]  | 70    | 50      | 951         |
| 121       | 8      | 6401       | 6988      | gi 1107528      | ctg start [Campylobacter coli]   | 70    | 45      | 588         |
| 131       | 8      | 6842       | 7936      | gi 1150454      | protease PpQ [Lactobacillus delbrueckii]   | 70    | 48      | 1095        |

TABLE 2

5. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match Accession | Match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 135       | 1      | 2          | 1489      | gi 3111309      | putative membrane-bound protein with four times repetition of ro-Ser-Ala at the N-terminus; function unknown (Alcaligenes utrophus)               | 70    | 49      | 1488        |
| 138       | 3      | 418        | 714       | gi 504181       | hypothetical protein (Bacillus subtilis)  | 70    | 46      | 297         |
| 164       | 8      | 9344       | 9874      | gi 49315        | ORF1 gene product (Bacillus subtilis)   | 70    | 47      | 531         |
| 164       | 16     | 115626     | 116618    | gi 1205212      | hypothetical protein (GB:D10483.18) (Haemophilus influenzae)  | 70    | 50      | 993         |
| 205       | 2      | 2735       | 1803      | gi 1215695      | peptide transport system protein SapF homolog; SapF homolog (Mycobacterium tuberculosis)  | 70    | 47      | 933         |
| 209       | 3      | 910        | 1386      | gi 1204665      | hypothetical protein (GB:X73124.26) (Haemophilus influenzae)  | 70    | 48      | 477         |
| 246       | 3      | 340        | 756       | gi 215098       | excisionase (Bacteriophage 154a)  | 70    | 46      | 417         |
| 263       | 7      | 7876       | 6749      | gi 142540       | aspartokinase II (Bacillus sp.)   | 70    | 51      | 1128        |
| 268       | 3      | 3212       | 4117      | gi 1340128      | ORF1 (Staphylococcus aureus)  | 70    | 50      | 906         |
| 302       | 6      | 3201       | 3827      | gi 147782       | frvA protein (gtg start) (Escherichia coli)   | 70    | 46      | 627         |
| 302       | 10     | 5879       | 7051      | pir C38530/C385 | guanine tRNA-ribosyltransferase (EC 2.4.2.29) - Escherichia coli  | 70    | 55      | 1173        |
| 313       | 1      | 2520       | 1414      | gi 1205934      | aminopeptidase a/I (Haemophilus influenzae)   | 70    | 46      | 1107        |
| 335       | 2      | 379        | 669       | gi 1070013      | protein-dependent (Bacillus subtilis)   | 70    | 48      | 291         |
| 403       | 1      | 1255       | 629       | gi 733147       | GuaF (Xanthomonas campestris)   | 70    | 33      | 627         |
| 444       | 10     | 4770       | 9273      | gi 1204742      | high affinity ribose transport protein (Haemophilus influenzae)   | 70    | 52      | 503         |
| 449       | 1      | 2          | 1243      | gi 619724       | MgtE (Bacillus firmus)  | 70    | 44      | 1242        |
| 472       | 1      | 637        | 320       | gi 727145       | open reading frame; putative (Bacillus amyloquelacene) pir a29091 a29091 hypothetical protein (bglA region) - Bacillus myoliquefaciens (fragment) | 70    | 41      | 318         |
| 480       | 2      | 727        | 1608      | gi 142560       | ATP synthase gamma subunit (Bacillus megaterium)  | 70    | 44      | 882         |
| 524       | 1      | 2          | 307       | gi 602292       | KCM2 protein (Brassica napus)   | 70    | 45      | 306         |
| 525       | 1      | 823        | 413       | gi 143372       | phosphoribosyl glycinamide formyltransferase (PUR-N) (Bacillus subtilis)  | 70    | 52      | 411         |
| 565       | 4      | 3625       | 2552      | gi 881434       | ORF1 (Bacillus subtilis)  | 70    | 51      | 1074        |
| 607       | 4      | 829        | 1284      | gi 511524       | hypothetical protein (SP:P37002) (Methanococcus jannaschii)   | 70    | 50      | 456         |
| 633       | 1      | 1383       | 703       | gi 431231       | luciferase (Bacillus caldolyticus)  | 70    | 53      | 681         |
| 646       | 3      | 1683       | 1309      | gi 467340       | unknown (Bacillus subtilis)   | 70    | 49      | 375         |
| 663       | 1      | 830        | 417       | gi 1303873      | V-lyz (Bacillus subtilis)   | 70    | 40      | 414         |

TABLE 2

saurus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 681       | 1      | 1488       | 781       | gi 4001678      | hypothetical protein [Synecocystis sp.]   | 70    | 53      | 708         |
| 708       | 1      | 2          | 448       | sp P3940 YOUN   | HYPOHETICAL 54.3 KD PROTEIN IN ECO-AKB INTERGENIC REGION.   | 70    | 51      | 447         |
| 725       | 1      | 51         | 722       | gi 1001644      | hypothetical protein [Synecocystis sp.]   | 70    | 48      | 672         |
| 776       | 1      | 1371       | 767       | gi 145165       | putative [Escherichia coli]   | 70    | 47      | 585         |
| 834       | 1      | 250        | 783       | gi 552971       | NADH dehydrogenase (ndhF) [Vicia faba]  | 70    | 47      | 534         |
| 865       | 2      | 1585       | 1379      | gi 1204636      | ATP-dependent helicase [Haemophilus influenzae]   | 70    | 45      | 207         |
| 894       | 1      | 535        | 269       | gi 467364       | DNA binding protein (probable) [Bacillus subtilis]  | 70    | 41      | 267         |
| 919       | 1      | 3          | 317       | gi 3314847      | ClnA [Bacillus subtilis]  | 70    | 40      | 315         |
| 944       | 1      | 3          | 572       | gi 709991       | hypothetical protein [Bacillus subtilis]  | 70    | 44      | 570         |
| 988       | 2      | 772        | 605       | gi 142441       | ORF 3: putative [Bacillus subtilis]   | 70    | 50      | 168         |
| 1055      | 1      | 3          | 335       | gi 529755       | laneC [Streptococcus pyogenes]  | 70    | 37      | 333         |
| 1093      | 1      | 2          | 904       | gi 853754       | ABC transporter [Bacillus subtilis]   | 70    | 49      | 903         |
| 1109      | 1      | 2          | 310       | gi 1001827      | hypothetical protein [Synecocystis sp.]   | 70    | 42      | 309         |
| 1220      | 1      | 468        | 235       | pir S23416 S234 | epib protein - Staphylococcus epidermidis   | 70    | 40      | 234         |
| 1279      | 1      | 73         | 348       | gi 153015       | FemA protein [Staphylococcus aureus]  | 70    | 47      | 276         |
| 1336      | 1      | 195        | 542       | sp P31776 PBPA  | PENICILLIN-BINDING PROTEIN 1A (PBP-1A) [PENICILLIN-BINDING PROTEIN A]                               | 70    | 50      | 348         |
| 1537      | 2      | 232        | 402       | gi 1148181      | putative [Bacillus subtilis]  | 70    | 50      | 171         |
| 1574      | 1      | 451        | 272       | gi 219630       | endothelin-A receptor [Homo sapiens]  | 70    | 47      | 180         |
| 1640      | 1      | 690        | 346       | gi 1146243      | 22.4% identity with Escherichia coli DNA-damage inducible protein ...; putative [Bacillus subtilis] | 70    | 46      | 345         |
| 2504      | 1      | 2          | 286       | gi 495179       | transmembrane protein [Lactococcus lactis]  | 70    | 51      | 285         |
| 3061      | 1      | 564        | 301       | gi 508175       | EITC domain of PTS-dependent Gat transport and phosphorylation Escherichia coli                     | 70    | 44      | 264         |
| 3128      | 1      | 2          | 199       | gi 1340096      | unknown [Mycobacterium tuberculosis]  | 70    | 51      | 198         |
| 3218      | 1      | 3          | 488       | gi 515938       | glutamate synthase (ferredoxin) [Synecocystis sp.] pir S46957 S46957                                | 70    | 50      | 486         |
| 3223      | 1      | 794        | 399       | gi 1154891      | ATP binding protein [Phormidium lewinosum]  | 70    | 52      | 396         |
| 3679      | 1      | 599        | 399       | gi 529385       | chromosome condensation protein [Caenorhabditis elegans]  | 70    | 30      | 201         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 3841      | 1      | 706        | 398       | gi1208965       | hypothetical 23.3 kd protein [Escherichia coli]   | 70    | 47      | 309         |
| 3929      | 1      | 3          | 401       | gi149435        | putative [Lactococcus lactis]   | 70    | 49      | 399         |
| 4044      | 1      | 595        | 374       | gi1602031       | similar to trimethylamine DH [Mycoplasma capricolus] pir[549950]549950<br>probable trimethylamine dehydrogenase (EC 5.99.7) - Mycoplasma capricolus (SGC3) (fragment) | 70    | 40      | 222         |
| 4329      | 1      | 558        | 280       | gi1339951       | small subunit of NADH-dependent glutamate synthase [Plectonoma boryanum]  | 70    | 49      | 279         |
| 4422      | 1      | 576        | 289       | gi296464        | ATPase [Lactococcus lactis]   | 70    | 57      | 288         |
| 4647      | 1      | 361        | 200       | gi166412        | NADH-glutamate synthase [Medicago sativa]   | 70    | 59      | 162         |
| 16        | 8      | 7571       | 9031      | gi1499620       | M. jannaschii predicted coding region M20798 [Methanococcus jannaschii]   | 69    | 44      | 1461        |
| 16        | 9      | 9080       | 10033     | gi1353197       | thioredoxin reductase [Eubacterium acidaminophilum]   | 69    | 54      | 954         |
| 30        | 1      | 1452       | 727       | gi1204910       | hypothetical protein (GB:U14003_302) [Haemophilus influenzae]   | 69    | 52      | 726         |
| 38        | 4      | 1023       | 1298      | gi1407773       | devA gene product [Anabaena sp.]  | 69    | 41      | 276         |
| 44        | 9      | 5987       | 6595      | gi1205920       | polyphosphate uptake system hydrophilic membrane-bound protein [Haemophilus influenzae]   | 69    | 45      | 609         |
| 62        | 15     | 9104       | 9475      | gi1385178       | unknown [Bacillus subtilis]   | 69    | 44      | 372         |
| 66        | 4      | 2402       | 2803      | gi1203893       | YopL [Bacillus subtilis]  | 69    | 51      | 402         |
| 67        | 15     | 14124      | 13627     | gi149647        | ORF2 [Listeria monocytogenes]   | 69    | 37      | 498         |
| 67        | 17     | 14053      | 14382     | gi1305002       | ORF_2356 [Escherichia coli]   | 69    | 49      | 310         |
| 67        | 19     | 15130      | 15807     | gi1105684       | ProV [Bacillus subtilis]  | 69    | 45      | 678         |
| 78        | 3      | 1447       | 2124      | gi1256633       | putative [Bacillus subtilis]  | 69    | 53      | 678         |
| 78        | 4      | 4513       | 3725      | gi1103958       | YqjG [Bacillus subtilis]  | 69    | 32      | 749         |
| 85        | 4      | 4521       | 4213      | pir1229326 E293 | hypothetical protein (pur operon) - Bacillus subtilis   | 69    | 32      | 309         |
| 86        | 6      | 3253       | 2854      | gi1473332       | orfC [Bacillus subtilis]  | 69    | 50      | 600         |
| 95        | 1      | 96         | 710       | gi1766468       | 4All antigen, sperm cell membrane antigen-putative sucrose-specific phosphotransferase enzyme II homolog (nlce, lctase, Peptide Partial, 72 aa)                       | 69    | 43      | 615         |
| 100       | 7      | 6023       | 7426      | gi1205355       | Na+/H+ antiporter [Haemophilus influenzae]  | 69    | 39      | 1404        |
| 102       | 2      | 2678       | 1650      | gi1561690       | glialycoprotease [Pasteurella haemolytica]  | 69    | 47      | 1029        |
| 103       | 8      | 112241     | 8537      | gi11009366      | Respiratory nitrate reductase [Bacillus subtilis]   | 69    | 54      | 3705        |
| 103       | 11     | 114987     | 12552     | gi1710020       | nitrite reductase (nirB) [Bacillus subtilis]  | 69    | 51      | 2436        |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession     | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|---------------------|---|-------|---------|-------------|
| 112       | 11     | 8708       | 10168     | [gi1354111]         | hexosephosphate transport protein [Salmonella typhimurium] p1r[041853] [041853] hexose phosphate transport system protein uhpF - salmonella typhimurium | 69    | 51      | 1461        |
| 112       | 16     | 16644      | 17414     | [gi1204435]         | pyruvate formate-lyase activating enzyme [Haemophilus influenzae]   | 69    | 50      | 771         |
| 113       | 2      | 33         | 953       | [gi1290509]         | o307 [Escherichia coli]   | 69    | 43      | 921         |
| 114       | 2      | 1537       | 1058      | [p1r142771] [1427]  | reticulocyte-binding protein 1 - Plasmodium vivax   | 69    | 39      | 480         |
| 121       | 6      | 4309       | 5310      | [gi1154633]         | hcrf [Bacillus subtilis]  | 69    | 53      | 1002        |
| 125       | 2      | 267        | 854       | [gi1413931]         | ipa-7d gene product [Bacillus subtilis]   | 69    | 43      | 588         |
| 149       | 27     | 110666     | 110400    | [p1r1528089] [5280] | hypothetical protein A - yeast (Zygosaccaromyces bisporus) plasmid pSU  | 69    | 39      | 267         |
| 161       | 1      | 1598       | 813       | [gi1205538]         | hypothetical protein (GB:U14001_2021) [Haemophilus influenzae]  | 69    | 47      | 786         |
| 165       | 4      | 2222       | 4633      | [gi140054]          | phenylalanyl-tRNA synthetase beta subunit (AA 1-804) [Bacillus bacillus]  | 69    | 52      | 2412        |
| 169       | 3      | 1210       | 1761      | [gi1256031]         | elongation factor Ts [Spirulina platensis]  | 69    | 45      | 552         |
| 175       | 12     | 8686       | 8339      | [gi1732682]         | F1ME protein [Escherichia coli]   | 69    | 69      | 348         |
| 190       | 2      | 484        | 1671      | [sp1p17731] [HIS8_  | HISTIDINOL-PHOSPHATE AMINOTRANSFERASE (EC 2.6.1.9) (IMIDAZOLE ACETOL-PHOSPHATE TRANSAMINASE)  | 69    | 48      | 1188        |
| 206       | 1      | 5551       | 2777      | [gi141750]          | hadr protein (AA 1-1033) [Escherichia coli]   | 69    | 49      | 2775        |
| 206       | 4      | 6038       | 5796      | [gi1256135]         | YhbF [Bacillus subtilis]  | 69    | 40      | 243         |
| 249       | 1      | 616        | 319       | [gi1405456]         | YncP [Bacillus subtilis]  | 69    | 50      | 1116        |
| 302       | 8      | 4820       | 5776      | [gi1001768]         | hypothetical protein [Synecococcus sp.]   | 69    | 48      | 957         |
| 324       | 2      | 7384       | 3893      | [gi1256798]         | pyruvate carboxylase [Rhizobium etli]   | 69    | 53      | 3492        |
| 351       | 3      | 2098       | 1808      | [gi1491664]         | T04H1.4 [Caenorhabditis elegans]  | 69    | 30      | 291         |
| 369       | 3      | 2075       | 2305      | [gi1316458]         | ORP [Balanophora acutrolata]  | 69    | 61      | 231         |
| 392       | 3      | 1999       | 2424      | [gi1556015]         | ORP [Bacillus subtilis]   | 69    | 45      | 426         |
| 410       | 1      | 87         | 779       | [gi1355611]         | phosphoglyceromutase [Zymomonas mobilis]  | 69    | 58      | 693         |
| 421       | 1      | 2085       | 1129      | [gi1276985]         | arginase [Bacillus caldovelox]  | 69    | 54      | 957         |
| 444       | 8      | 6713       | 7741      | [gi13221782]        | purine synthesis repressor [Haemophilus influenzae]   | 69    | 40      | 1029        |
| 453       | 1      | 828        | 415       | [gi13122758]        | unknown [Bacillus subtilis]   | 69    | 57      | 414         |
| 469       | 2      | 3286       | 2246      | [gi1458228]         | mutY homolog [Homo sapiens]   | 69    | 44      | 1041        |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name  | V sim | V ident | Length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 509       | 3      | 1730       | 1371      | gi149224        | USF 4 [Synecoccus sp.]   | 69    | 39      | 360         |
| 520       | 5      | 3023       | 2823      | gi1726427       | similar to D. melanogaster HST101-2 protein (PIR:S3154) Caenorhabditis elegans   | 69    | 39      | 201         |
| 531       | 1      | 26         | 760       | gi1509672       | repressor protein (bacteriophage Tuc2009)  | 69    | 33      | 735         |
| 589       | 1      | 107        | 253       | gi149101        | 17.9 kDa heat shock protein (hsp17.9) (P1aum sativum)  | 69    | 52      | 147         |
| 594       | 2      | 597        | 1391      | gi1742783       | DNA photolyase (Bacillus firmus)   | 69    | 48      | 795         |
| 604       | 4      | 2476       | 2114      | gi1413930       | lipa-6d gene product (Bacillus subtilis)   | 69    | 45      | 363         |
| 607       | 1      | 2          | 313       | gi1236103       | W0802.3 (Caenorhabditis elegans)   | 69    | 47      | 312         |
| 607       | 2      | 590        | 312       | gi1536715       | ORF YBR275c (Saccharomyces cerevisiae)   | 69    | 39      | 279         |
| 734       | 1      | 864        | 433       | gi1467327       | unknown (Bacillus subtilis)  | 69    | 44      | 432         |
| 759       | 1      | 3          | 338       | gi11009367      | Respiratory nitrate reductase (Bacillus subtilis)  | 69    | 50      | 336         |
| 761       | 2      | 392        | 586       | gi13508         | Leucyl-tRNA synthetase (cytoplasmic) (Saccharomyces cerevisiae) [ORF YPL160w (Saccharomyces cerevisiae)]                         | 69    | 46      | 195         |
| 802       | 1      | 72         | 1013      | gi1143044       | ferrochelatase (Bacillus subtilis)   | 69    | 55      | 942         |
| 816       | 1      | 2573       | 1368      | gi1510268       | restriction modification system S subunit (Methanococcus jannaschii)   | 69    | 45      | 1206        |
| 838       | 2      | 133        | 387       | gi11255371      | coded for by C. elegans cDNA YK349.5; coded for by C. elegans cDNA YK349.3; Similar to guanylate kinase (Caenorhabditis elegans) | 69    | 46      | 255         |
| 851       | 2      | 745        | 1005      | gi1288998       | IsocA gene product (Antilimonium sp.)  | 69    | 39      | 261         |
| 867       | 1      | 535        | 269       | gi11070014      | protein-dependent (Bacillus subtilis)  | 69    | 47      | 267         |
| 995       | 1      | 954        | 478       | gi1205569       | transcription elongation factor (Haemophilus influenzae)   | 69    | 53      | 477         |
| 999       | 1      | 1009       | 506       | gi1899254       | predicted trithorax protein (Drosophila virilis)   | 69    | 21      | 504         |
| 1127      | 1      | 1315       | 659       | gi1205434       | H. influenzae predicted coding region H1191 (Haemophilus influenzae)   | 69    | 56      | 657         |
| 1138      | 1      | 248        | 460       | gi1510646       | H. jannaschii predicted coding region HJ0568 (Methanococcus jannaschii)  | 69    | 48      | 213         |
| 2928      | 1      | 3          | 401       | gi1205003       | glutamate permease (Escherichia coli)  | 69    | 41      | 399         |
| 3090      | 1      | 444        | 223       | gi11204987      | DNA polymerase III, alpha chain (Haemophilus influenzae)   | 69    | 36      | 222         |
| 3817      | 1      | 2          | 400       | gi11483199      | peptide-synthetase (Amycolatopsis mediterranei)  | 69    | 45      | 399         |
| 3833      | 1      | 667        | 335       | gi1524193       | unknown (Mycobacterium tuberculosis)   | 69    | 46      | 333         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Accession        | Match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|------------------|--|-------|---------|-------------|
| 4079      | 1      | 747        | 400       | gi1546918        | orfY 3' of coxK [Bacillus subtilis, 526, Peptide Partial, 140 aa]<br>pir153612 [S1612 hypothetical protein Y - Bacillus subtilis<br>sp140398 [YHND_BACSU HYPOTHETICAL PROTEIN IN COXK 3' REGION (ORFY FRAGMENT)] | 69    | 64      | 348         |
| 4115      | 2      | 215        | 400       | gi1517205        | 167 kDa Myosin-crossreactive streptococcal antigen [Streptococcus yogenes]   | 69    | 59      | 186         |
| 4139      | 1      | 1          | 333       | gi11208451       | hypothetical protein [Synecocystis sp.]  | 69    | 36      | 333         |
| 4258      | 1      | 457        | 230       | gi1496158        | restriction-modification enzyme subunit M1 [Mycoplasma pulmonis]<br>pir1549395 [S49395 HsdM1 protein - Mycoplasma pulmonis (SOC)]  | 69    | 43      | 228         |
| 4317      | 1      | 90         | 374       | gi1413967        | lpa-43d gene product [Bacillus subtilis]   | 69    | 44      | 285         |
| 4465      | 1      | 3          | 293       | gi1396296        | similar to phosphotransferase system enzyme II [Escherichia coli]<br>sp1432672 [PTWC_ECOLI PTS SYSTEM, FRUCTOSE-LIKE-2 IIC COMPONENT<br>PHOSPHOTRANSFERASE ENZYME II, C COMPONENT]                               | 69    | 49      | 291         |
| 3         | 1      | 2302       | 1193      | gi11109685       | ProM [Bacillus subtilis]   | 68    | 46      | 1110        |
| 15        | 4      | 2592       | 2074      | gi1807973        | unknown [Saccharomyces cerevisiae]   | 68    | 45      | 519         |
| 31        | 8      | 6328       | 8772      | gi1290642        | ATPase [Enterococcus hirae]  | 68    | 48      | 2445        |
| 40        | 2      | 1115       | 750       | gi1606342        | ORF_0622; reading frame open far upstream of start; possible transmitt. linking to previous ORF [Escherichia coli]   | 68    | 55      | 166         |
| 46        | 9      | 6866       | 8415      | gi1155276        | aldehyde dehydrogenase [Vibrio cholerae]   | 68    | 44      | 1530        |
| 48        | 3      | 3643       | 3404      | gi1285608        | 241k polyprotein [Apple stem grooving virus]   | 68    | 47      | 240         |
| 48        | 4      | 3536       | 4132      | gi11045937       | M. genitalium predicted coding region HQ246 [Mycoplasma genitalium]  | 68    | 39      | 597         |
| 53        | 10     | 11671      | 10685     | gi11303952       | YqJA [Bacillus subtilis]   | 68    | 46      | 987         |
| 70        | 9      | 7346       | 8155      | gi1147198        | phnE protein [Escherichia coli]  | 68    | 40      | 810         |
| 89        | 4      | 1899       | 2966      | gi1145173        | 35 kDa protein [Escherichia coli]  | 68    | 43      | 1068        |
| 108       | 3      | 2187       | 1150      | gi138722         | precursor (aa -20 to 381) [Acinetobacter calcoaceticus] ir1A29277 [A29277<br>aldose 1-epimerase (EC 5.1.3.3) - Acinetobacter calcoaceticus]  | 68    | 57      | 1038        |
| 112       | 5      | 2666       | 3622      | gi1153724        | MalC [Streptococcus pneumoniae]  | 68    | 55      | 957         |
| 116       | 7      | 7865       | 8638      | gi1143608        | sporulation protein [Bacillus subtilis]  | 68    | 48      | 774         |
| 118       | 3      | 2484       | 3698      | gi11303805       | Yqer [Bacillus subtilis]   | 68    | 46      | 1215        |
| 120       | 2      | 1624       | 1594      | sp1438038 [CYSJ_ | SULFITE REDUCTASE (NADPH) FLAVOPROTEIN ALPHA-COMPONENT (EC 1.8.1.2) (SIR-<br>FP)]  | 68    | 45      | 171         |
| 129       | 1      | 1          | 1011      | gi1396307        | argininosuccinate lyase [Escherichia coli]   | 68    | 50      | 1011        |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name  | % sim | % ident | Length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 132       | 3      | 1867       | 2739      | gi 1516267      | ORF2 [Bacillus megaterium]   | 68    | 48      | 873         |
| 134       | 2      | 848        | 1012      | gi 147545       | DNA recombinase [Escherichia coli]                                       | 68    | 50      | 165         |
| 141       | 2      | 372        | 614       | gi 972116       | act (stress inducible protein) [Glycine max]                             | 68    | 36      | 243         |
| 149       | 7      | 2454       | 2260      | gi 145774       | hsp70 protein (dnaK gene) [Escherichia coli]                             | 68    | 48      | 195         |
| 155       | 2      | 1776       | 1534      | gi 216583       | ORF1 [Escherichia coli]  | 68    | 38      | 243         |
| 158       | 3      | 1826       | 3289      | gi 933940       | HYDROMETICAL 56.3 KD PROTEIN IN ECO-ALKB INTERGENIC REGION.              | 68    | 51      | 1464        |
| 169       | 6      | 2749       | 3318      | gi 1403402      | unknown [Mycobacterium tuberculosis]                                     | 68    | 46      | 570         |
| 175       | 10     | 9158       | 7365      | gi 1072195      | phaA gene product [Rhizobium moliotii]                                   | 68    | 51      | 1794        |
| 188       | 7      | 4184       | 5434      | gi 1173843      | 3-ketoacyl-ACP synthase II [Vibrio Harveyi]                              | 68    | 48      | 1251        |
| 189       | 3      | 907        | 1665      | gi 1467383      | DNA binding protein (probable) [Bacillus subtilis]                       | 68    | 55      | 759         |
| 206       | 5      | 7683       | 6709      | gi 1256138      | YbbI [Bacillus subtilis]   | 68    | 48      | 975         |
| 206       | 8      | 110425     | 12176     | gi 1452687      | pyruvate decarboxylase [Saccharomyces cerevisiae]                        | 68    | 48      | 1752        |
| 212       | 8      | 3421       | 1648      | gi 1369941      | cl gene product [Bacteriophage B1]                                       | 68    | 39      | 228         |
| 214       | 8      | 5457       | 6482      | gi 1420467      | ORF YOR196c [Saccharomyces cerevisiae]                                   | 68    | 45      | 1026        |
| 237       | 4      | 2507       | 3088      | gi 149381       | HLH [Lactococcus lactis]   | 68    | 46      | 582         |
| 243       | 5      | 5540       | 4542      | gi 1235684      | mevalonate pyrophosphate decarboxylase [Saccharomyces cerevisiae]        | 68    | 47      | 999         |
| 262       | 1      | 3          | 164       | gi 150974       | 4-oxalocrotonate tautomerase [Pseudomonas putida]                        | 68    | 42      | 162         |
| 262       | 2      | 1984       | 1118      | gi 1147744      | PSR [Enterococcus hirae]   | 68    | 49      | 867         |
| 276       | 6      | 3702       | 3139      | gi 930750       | ATP-BINDING PROTEIN ABC (FRAGMENT)                                       | 68    | 50      | 564         |
| 306       | 6      | 6345       | 5725      | gi 1256617      | adenine phosphoribosyltransferase [Bacillus subtilis]                    | 68    | 53      | 621         |
| 333       | 3      | 4599       | 3850      | gi 1467473      | unknown [Bacillus subtilis]  | 68    | 45      | 750         |
| 365       | 6      | 5017       | 4838      | gi 1130643      | T2283.3 [Caenorhabditis elegans]   | 68    | 45      | 180         |
| 376       | 2      | 549        | 1646      | gi 1277026      | DAPA aminotransferase [Bacillus subtilis]                                | 68    | 51      | 1094        |
| 405       | 1      | 1741       | 872       | gi 1302917      | YqjB [Bacillus subtilis]   | 68    | 47      | 870         |
| 406       | 2      | 853        | 539       | gi 1511513      | ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii] | 68    | 44      | 315         |
| 426       | 6      | 3558       | 3391      | gi 624632       | glcL [Escherichia coli]  | 68    | 48      | 168         |
| 438       | 1      | 108        | 329       | gi 146923       | nitrogenase reductase [Escherichia coli]                                 | 68    | 43      | 222         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 443       | 1      | 476        | 240       | gi 535810       | hippuricase [Campylobacter jejuni]   | 68    | 42      | 237         |
| 443       | 2      | 518        | 1015      | gi 1204742      | H. influenzae predicted coding region H10491 [Haemophilus influenzae]  | 68    | 48      | 498         |
| 443       | 5      | 4447       | 3779      | gi gn-660       | deoxyribose-phosphate aldolase [Bacillus subtilis] pif[S49455]S49455   | 68    | 55      | 669         |
| 476       | 2      | 240        | 1184      | gi 571345       | unknown, similar to E.coli cardiolipin synthase [Bacillus subtilis] sp P45860 YME_BACSU HYPOTHETICAL 58.2 PROTEIN IN NARI-ACDA NTERGENIC REGION.               | 68    | 45      | 945         |
| 486       | 2      | 1876       | 1046      | gi 147328       | transport protein [Escherichia coli]   | 68    | 41      | 831         |
| 517       | 3      | 1764       | 2084      | gi 523809       | orf2 [Bacteriophage A2]  | 68    | 64      | 321         |
| 572       | 1      | 2          | 571       | sp P19237 Y03L  | HYPOTHETICAL 6.8 KD PROTEIN IN MRDC-TX INTERGENIC REGION.  | 68    | 47      | 570         |
| 646       | 1      | 914        | 459       | gi 413982       | ipa-58r gene product [Bacillus subtilis]   | 68    | 52      | 456         |
| 659       | 3      | 1668       | 1901      | gi 1107541      | C3109.8 [Caenorhabditis elegans]   | 68    | 36      | 234         |
| 864       | 5      | 1510       | 1716      | gi 1145774      | hsp70 protein [dnaK gene] [Escherichia coli]   | 68    | 48      | 207         |
| 920       | 1      | 860        | 432       | gi 1110416      | Hypothetical protein (SP-P31466) [Methanococcus jennaschii]  | 68    | 54      | 429         |
| 952       | 1      | 1096       | 611       | gi C33456       | reductase [Leishmania major]   | 68    | 46      | 486         |
| 970       | 1      | 91         | 402       | gi 11354775     | pfos/R [Treponema pallidum]  | 68    | 46      | 312         |
| 1028      | 1      | 1064       | 534       | gi 1110117      | glutaminylate decarboxylase [Bacillus subtilis]  | 68    | 47      | 531         |
| 1029      | 1      | 428        | 216       | gi 1135714      | Plasmodium falciparum mRNA for asparagine-rich antigen (clone 17c1) [Plasmodium falciparum]  | 68    | 31      | 213         |
| 1058      | 1      | 692        | 348       | gi 581649       | epic gene product [Staphylococcus epidermidis]   | 68    | 46      | 345         |
| 1056      | 2      | 665        | 465       | gi 143434       | Rho Factor [Bacillus subtilis]   | 68    | 43      | 201         |
| 1308      | 1      | 2          | 694       | gi 1169939      | group B oligopeptidase.PepB [Streptococcus agalactiae]   | 68    | 50      | 693         |
| 1679      | 1      | 2          | 238       | gi 117205       | 67 kDa Myosin-crossreactive streptococcal antigen [Streptococcus yogenes]  | 68    | 53      | 237         |
| 2039      | 1      | 3          | 383       | gi 153898       | transport protein [Salmonella typhimurium]   | 68    | 51      | 381         |
| 2077      | 1      | 3          | 326       | pir C33496 C334 | hisc homolog - Bacillus subtilis   | 68    | 47      | 324         |
| 2112      | 1      | 613        | 374       | gi 64884        | lamin LII [Xenopus laevis]   | 68    | 50      | 240         |
| 2273      | 1      | 793        | 398       | gi 581648       | epib gene product [Staphylococcus epidermidis]   | 68    | 45      | 396         |
| 2948      | 1      | 2          | 385       | gi 216869       | branched-chain amino acid transport carrier [Pseudomonas aeruginosa] pir A38534 A38534 branched-chain amino acid transport protein braz Pseudomonas aeruginosa | 68    | 41      | 384         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 2955      | 1      | 768        | 400       | gi 904179       | hypothetical protein (Bacillus subtilis)   | 68    | 49      | 369         |
| 2981      | 1      | 572        | 288       | gi 508979       | CTP-binding protein (Bacillus subtilis)  | 68    | 48      | 285         |
| 3014      | 1      | 584        | 294       | gi 1524194      | ORF-2 upstream of gsaB operon (Bacillus subtilis)  | 68    | 45      | 291         |
| 3082      | 1      | 336        | 169       | gi 1204696      | fructose-permease IBC component (Haemophilus influenzae)   | 68    | 53      | 168         |
| 3108      | 1      | 103        | 258       | gi 217855       | heat-shock protein (Xanthomonas thaliana)  | 68    | 48      | 156         |
| 3639      | 1      | 919        | 461       | gi 1510490      | nitrate transport permease protein (Methanococcus jannaschii)  | 68    | 47      | 459         |
| 3657      | 1      | 1          | 330       | gi 155169       | PFS enzyme-II fructose (Xanthomonas campestris)  | 68    | 48      | 330         |
| 3823      | 1      | 780        | 391       | gi 603768       | HucU protein, imidazole-5-propionate hydrolase (Bacillus subtilis)   | 68    | 54      | 390         |
| 3982      | 1      | 2          | 277       | gi 149435       | putative (Lactococcus lactis)  | 68    | 47      | 276         |
| 4051      | 1      | 1          | 342       | gi 450888       | hadM gene of Ecoprr gene product (Escherichia coli) pir S38437 S38437 hadM protein - Escherichia coli pir S09629 S09629 hypothetical protein A - Escherichia coli (SUB 40-520) | 68    | 48      | 342         |
| 4089      | 1      | 12         | 209       | gi 1353678      | heavy-metal transporting P-type ATPase (Proteus mirabilis)   | 68    | 47      | 198         |
| 4143      | 1      | 47         | 187       | gi 603769       | HucU protein, urocanase (Bacillus subtilis)  | 68    | 55      | 141         |
| 4148      | 1      | 2          | 352       | gi 450688       | hadM gene of Ecoprr gene product (Escherichia coli) pir S38437 S38437 hadM protein - Escherichia coli pir S09629 S09629 hypothetical protein A - Escherichia coli (SUB 40-520) | 68    | 51      | 351         |
| 4173      | 1      | 2          | 382       | gi 1041097      | Pyruvate Kinase (Bacillus psychrophilus)   | 68    | 48      | 381         |
| 4182      | 1      | 498        | 250       | gi 413968       | lpa-44d gene product (Bacillus subtilis)   | 68    | 50      | 249         |
| 4362      | 2      | 148        | 318       | gi 450688       | hadM gene of Ecoprr gene product (Escherichia coli) pir S38437 S38437 hadM protein - Escherichia coli pir S09629 S09629 hypothetical protein A - Escherichia coli (SUB 40-520) | 68    | 44      | 171         |
| 5         | 11     | 9493       | 8300      | gi 1143727      | putative (Bacillus subtilis)   | 67    | 46      | 1194        |
| 31        | 11     | 10318      | 9833      | gi 216746       | D-lactate dehydrogenase (Lactobacillus plantarum)  | 67    | 41      | 486         |
| 32        | 3      | 1560       | 3155      | gi 1098557      | renal sodium/dicarboxylate cotransporter (Homo sapiens)  | 67    | 46      | 1596        |
| 32        | 5      | 4945       | 4165      | gi 1510720      | prephenate dehydratase (Methanococcus jannaschii)  | 67    | 51      | 801         |
| 36        | 5      | 5350       | 4268      | gi 146216       | 45% identity with the product of the ORF6 gene from the Erwinia herbicola carotenoid biosynthesis cluster; putative (Bacillus subtilis)  | 67    | 58      | 1083        |
| 44        | 7      | 4492       | 5304      | gi 1006621      | hypothetical protein (Synecocystis sp.)  | 67    | 43      | 813         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 56        | 7      | 3943       | 8481      | gi 304131       | glutamate synthase large subunit precursor (Acetivibrium brasilense) p1r B46602 B46602 glutamate synthase (NADPH) (EC 1.4.1.13) alpha chain - Acetivibrium brasilense | 67    | 52      | 4539        |
| 56        | 12     | 13923      | 14678     | gi 1000453      | TrpR (Bacillus subtilis)  | 67    | 48      | 756         |
| 62        | 8      | 5092       | 4757      | gi 1113949      | orf3 (Bacillus, C-125, alkali-sensitive mutant 18224, Peptide Mutant, 112 aa)   | 67    | 45      | 336         |
| 62        | 10     | 7570       | 6338      | gi 654655       | Na/H antiporter system (Bacillus alcalophilus)  | 67    | 49      | 1233        |
| 99        | 3      | 2119       | 3321      | gi 1204349      | hypothetical protein (GB:D90212.3) (Haemophilus influenzae)   | 67    | 50      | 1203        |
| 102       | 9      | 5695       | 7176      | gi 149432       | putative (Lactococcus lactis)   | 67    | 51      | 1482        |
| 103       | 13     | 14549      | 14049     | gi 1408497      | LPD gene product (Bacillus subtilis)  | 67    | 48      | 501         |
| 109       | 15     | 14821      | 13982     | gi 413976       | ipa-52r gene product (Bacillus subtilis)  | 67    | 49      | 840         |
| 109       | 17     | 14811      | 15194     | gi 413983       | ipa-59d gene product (Bacillus subtilis)  | 67    | 29      | 384         |
| 121       | 4      | 1713       | 2153      | gi 1262335      | YnaA (Bacillus subtilis)  | 67    | 54      | 441         |
| 122       | 3      | 1          | 1149      | gi 143047       | ORF8 (Bacillus subtilis)  | 67    | 35      | 1149        |
| 124       | 5      | 4060       | 3518      | gi 556885       | unknown (Bacillus subtilis)   | 67    | 47      | 543         |
| 131       | 2      | 4584       | 3589      | gi 1046081      | hypothetical protein (GB:D26145.10) (Mycoplasma genitalium)   | 67    | 30      | 996         |
| 140       | 3      | 2499       | 2297      | gi 146549       | kdsC (Escherichia coli)   | 67    | 45      | 603         |
| 142       | 4      | 5409       | 4198      | gi 1212775      | GTP cyclohydrolase II (Bacillus amyloquelicensis)   | 67    | 55      | 1212        |
| 147       | 5      | 2913       | 2374      | gi 1303709      | YrA3 (Bacillus subtilis)  | 67    | 44      | 540         |
| 152       | 8      | 6341       | 6672      | gi 1377841      | unknown (Bacillus subtilis)   | 67    | 46      | 333         |
| 161       | 4      | 2720       | 3763      | gi 496319       | SpxB (Synechococcus sp.)  | 67    | 47      | 1044        |
| 163       | 6      | 1989       | 3428      | gi 595681       | 2-oxoglutarate/malate translocator (Spinacia oleracea)  | 67    | 47      | 1440        |
| 193       | 3      | 1351       | 1626      | gi 1511101      | shikimate 5-dehydrogenase (Methanococcus jannaschii)  | 67    | 53      | 276         |
| 200       | 2      | 917        | 2179      | gi 142439       | ATP-dependent nuclease (Bacillus subtilis)  | 67    | 48      | 1263        |
| 206       | 10     | 12445      | 12401     | sp P37347 VECD  | HYPOTHEICAL 21.8 KD PROTEIN IN ASPS 5'-REGION   | 67    | 47      | 357         |
| 206       | 11     | 13047      | 14432     | gi 732813       | branched-chain amino acid carrier (Lactobacillus delbrueckii)   | 67    | 46      | 1386        |
| 208       | 2      | 1321       | 809       | gi 1033037      | 100 kDa heat shock protein (Hsp100) (Leishmania major)  | 67    | 36      | 513         |
| 238       | 3      | 1019       | 2052      | gi 809542       | CbrA protein (Erwinia chrysanthemi)   | 67    | 42      | 1014        |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 246       | 2      | 176        | 367       | gi 1215098      | lecA [Bacteriophage 154a]   | 67    | 37      | 192         |
| 276       | 2      | 2260       | 1412      | gi 1303560      | ORF271 [Escherichia coli]   | 67    | 50      | 849         |
| 297       | 6      | 2223       | 3056      | gi 1142784      | lcaA protein [Bacillus firmus]  | 67    | 46      | 834         |
| 307       | 7      | 5220       | 4186      | gi 1070013      | protein-dependent [Bacillus subtilis]   | 67    | 43      | 1035        |
| 316       | 1      | 36         | 1028      | gi 1161061      | dioloxigenase [Methylobacterium extorquens]   | 67    | 52      | 993         |
| 324       | 3      | 5650       | 5030      | gi 11469784     | putative cell division protein ftsW [Enterococcus hirae]  | 67    | 49      | 621         |
| 336       | 1      | 524        | 264       | gi 173122       | urea amidolyase [Saccharomyces cerevisiae]  | 67    | 45      | 261         |
| 360       | 1      | 108        | 1394      | gi 1305315711   | 11S121DYL-THNA SYNTHETASE (EC 6.1.1.21) [HISTIDINE--THNA LIGASE] (HISMG)  | 67    | 47      | 1287        |
| 364       | 3      | 4890       | 3592      | gi 1151259      | HM-CoA reductase (EC 1.1.1.88) [Pseudomonas nevalonii] p1r[A44756]A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.      | 67    | 46      | 1299        |
| 365       | 3      | 2940       | 2113      | gi 1296823      | lorf2 gene product [Lactobacillus helveticus]   | 67    | 47      | 828         |
| 367       | 2      | 325        | 918       | gi 1039479      | ORF1 [Lactococcus lactis]   | 67    | 47      | 594         |
| 395       | 3      | 666        | 1271      | gi 1204516      | hypothetical protein (GI:U00014_4) [Haemophilus influenzae]   | 67    | 55      | 606         |
| 415       | 1      | 1800       | 901       | gi 1882579      | CG Site No. 29739 [Escherichia coli]  | 67    | 46      | 900         |
| 419       | 1      | 1799       | 903       | gi 520752       | putative [Bacillus subtilis]  | 67    | 46      | 897         |
| 474       | 1      | 2          | 796       | gi 486906       | argininosuccinate synthetase [Streptomyces clavuligerus] p1r[S57659]S57659 argininosuccinate synthetase (EC 6.3.4.5) - treptomyces clavuligerus | 67    | 49      | 794         |
| 485       | 2      | 1921       | 2226      | gi 143434       | Rho factor [Bacillus subtilis]  | 67    | 43      | 306         |
| 596       | 1      | 1728       | 865       | gi 1303853      | Yggf [Bacillus subtilis]  | 67    | 47      | 864         |
| 700       | 1      | 433        | 218       | gi 1204628      | hypothetical protein (SP-E21498) [Haemophilus influenzae]   | 67    | 47      | 216         |
| 806       | 2      | 249        | 647       | gi 677947       | AppC [Bacillus subtilis]  | 67    | 51      | 399         |
| 828       | 2      | 340        | 900       | gi 777761       | lraA [Synchococcus sp.]   | 67    | 37      | 561         |
| 833       | 1      | 1407       | 916       | gi 142996       | regulatory protein [Bacillus subtilis]  | 67    | 41      | 492         |
| 856       | 1      | 1555       | 779       | gi 780224       | 2K970.2 [Caenorhabditis elegans]  | 67    | 38      | 777         |
| 888       | 1      | 1614       | 850       | gi 137315       | TTG start codon [Bacillus licheniformis]  | 67    | 40      | 765         |
| 1034      | 1      | 1190       | 597       | gi 1205113      | hypothetical protein (GI:U9201_15) [Haemophilus influenzae]   | 67    | 45      | 594         |
| 1062      | 1      | 636        | 319       | gi 1303850      | YggC [Bacillus subtilis]  | 67    | 41      | 318         |
| 1067      | 1      | 918        | 460       | pir[AJ2950]AJ29 | probable reductase protein - Leishmania major   | 67    | 54      | 459         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 1358      | 1      | 3          | 293       | gi11001369      | hypothetical protein [Synchocystis sp.]   | 67    | 44      | 291         |
| 2181      | 1      | 3          | 302       | gi11510416      | hypothetical protein (SF-P1466) [Methanococcus jannaschii]  | 67    | 48      | 300         |
| 3000      | 1      | 1          | 507       | gi1517205       | 67 kDa Myosin-crossreactive streptococcal antigen [Streptococcus pyogenes]  | 67    | 56      | 507         |
| 3066      | 1      | 464        | 234       | gi1308861       | OTG start codon [Lactococcus lactis]  | 67    | 46      | 231         |
| 3087      | 1      | 454        | 251       | gi11205366      | oligopeptide transport ATP-binding protein [Haemophilus influenzae]   | 67    | 41      | 204         |
| 3101      | 1      | 2          | 256       | gi13533541      | uroporphyrinogen III methyltransferase [Zea mays]   | 67    | 55      | 255         |
| 3598      | 1      | 728        | 293       | gi1151259       | HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas mvalonii] pir[A44756] A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp. | 67    | 56      | 336         |
| 3765      | 2      | 584        | 366       | gi1557489       | menD [Bacillus subtilis]  | 67    | 45      | 219         |
| 3788      | 1      | 658        | 398       | pir1552915 5529 | nitrate reductase alpha chain - Bacillus subtilis (fragment)  | 67    | 45      | 261         |
| 3883      | 1      | 2          | 285       | gi1704397       | cystathionine beta-lyase [Arabidopsis thaliana]   | 67    | 46      | 284         |
| 3926      | 1      | 2          | 340       | gi11463199      | peptide-synthetase [Amycolatopsis mediterranea]   | 67    | 44      | 339         |
| 4417      | 1      | 82         | 396       | gi11205337      | ribonucleotide transport ATP-binding protein [Haemophilus influenzae]   | 67    | 46      | 315         |
| 2         | 3      | 3075       | 389       | gi1535348       | CodV [Bacillus subtilis]  | 66    | 42      | 915         |
| 15        | 6      | 2273       | 2542      | gi146491        | SmcB [Synchococcus PCC7942]   | 66    | 37      | 270         |
| 11        | 9      | 8059       | 7826      | gi1292046       | luciferin [Momo sapiens]  | 66    | 44      | 234         |
| 31        | 10     | 9034       | 9258      | gi11204545      | mercury scavenger protein [Haemophilus influenzae]  | 66    | 48      | 225         |
| 32        | 6      | 6347       | 3253      | gi1598342       | inducible nitric oxide synthase [Gallus gallus]   | 66    | 47      | 1095        |
| 44        | 13     | 8856       | 10124     | gi11510751      | molybdenum cofactor biosynthesis moeA protein [Methanococcus jannaschii]  | 66    | 46      | 1269        |
| 48        | 2      | 1276       | 2868      | gi1150209       | ORF 1 [Mycoplasma mycoides]   | 66    | 40      | 1593        |
| 58        | 8      | 7178       | 8428      | gi1665999       | hypothetical protein [Bacillus subtilis]  | 66    | 47      | 1251        |
| 62        | 7      | 5143       | 4370      | gi11072398      | phd gene product [Rhizobium meliloti]   | 66    | 40      | 774         |
| 70        | 14     | 11693      | 10998     | gi1809660       | deoxyribose-phosphate aldolase [Bacillus subtilis] pir[S49455] S49455 deoxyribose-phosphate aldolase (EC 4.1.2.4) - acillus subtilis        | 66    | 55      | 696         |
| 76        | 1      | 1          | 1305      | gi1142440       | ATP-dependent nuclease [Bacillus subtilis]  | 66    | 42      | 1305        |
| 91        | 6      | 9236       | 8205      | gi1704397       | cystathionine beta-lyase [Arabidopsis thaliana]   | 66    | 43      | 1032        |
| 102       | 5      | 3810       | 2265      | gi11204323      | hypothetical protein (SF-P1805) [Haemophilus influenzae]  | 66    | 41      | 546         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 103       | 4      | 3418       | 2732      | gi 971344       | nitrate reductase gamma subunit [Bacillus subtilis] sp P42177 NARI_BACSU NITRATE REDUCTASE GAMMA SUBUNIT (EC 1.7.99.4) - gi 1009369 Respiratory nitrate reductase [Bacillus subtilis] (SUB -160) | 66    | 48      | 687         |
| 109       | 6      | 4243       | 4674      | gi 170886       | glucosamine-6-phosphate deaminase [Candida albicans] pI A46652 A46652 glucosamine-6-phosphate isomerase (EC 5.3.1.10) - east [Candida albicans]  | 66    | 45      | 432         |
| 112       | 17     | 17491      | 17712     | gi 1323179      | ORF YGR111w [Saccharomyces cerevisiae]   | 66    | 33      | 222         |
| 116       | 2      | 4667       | 2637      | gi 1491813      | gamma-glutamyltranspeptidase [Bacillus subtilis]   | 66    | 43      | 2011        |
| 150       | 5      | 3189       | 2989      | gi 1146224      | putative [Bacillus subtilis]   | 66    | 30      | 201         |
| 172       | 5      | 3264       | 3662      | gi 755152       | highly hydrophobic integral membrane protein [Bacillus subtilis] sp P42953 TAGC_BACSU TEICHOIC ACID TRANSLLOCATION PERMEASE PROTEIN AGG.   | 66    | 41      | 399         |
| 174       | 5      | 4592       | 3723      | gi 1146241      | pantothenate synthetase [Bacillus subtilis]  | 66    | 49      | 870         |
| 175       | 4      | 3209       | 2880      | gi 642655       | unknown [Rhizobium meliloti]   | 66    | 29      | 310         |
| 175       | 11     | 8743       | 7994      | gi 854655       | Na/H antiporter system [Bacillus alcalophilus]   | 66    | 43      | 750         |
| 190       | 5      | 7079       | 5727      | gi 451072       | di-tripeptide transporter [Lactococcus lactis]   | 66    | 40      | 1353        |
| 195       | 15     | 113919     | 11713     | gi 1322411      | unknown [Mycobacterium tuberculosis]   | 66    | 42      | 207         |
| 217       | 3      | 2822       | 2595      | gi 1141542      | alternative stop codon [Rattus norvegicus]   | 66    | 36      | 228         |
| 233       | 9      | 7133       | 6135      | gi 1458327      | P08P3.4 gene product [Caenorhabditis elegans]  | 66    | 47      | 999         |
| 238       | 1      | 41         | 1041      | gi 809541       | CdrA protein [Brwnia chrysanthemi]   | 66    | 42      | 994         |
| 241       | 1      | 2102       | 1053      | gi 153067       | peptidoglycan hydrolase [Staphylococcus aureus]  | 66    | 53      | 1050        |
| 261       | 1      | 1178       | 648       | gi 1510859      | M. jannaschii predicted coding region M0790 [Methanococcus jannaschii]   | 66    | 40      | 531         |
| 263       | 3      | 3731       | 2973      | gi 1205865      | tetrahydrodipicolinate N-succinyltransferase [Haemophilus influenzae]  | 66    | 47      | 759         |
| 272       | 8      | 6548       | 5484      | gi 882101       | high affinity nickel transporter [Alcaligenes eutrophus] sp P23516 HUXM_ALCEU HIGH-AFFINITY NICKEL TRANSPORT PROTEIN.  | 66    | 44      | 1065        |
| 276       | 3      | 2805       | 2104      | gi 1208965      | hypothetical 23.3 kd protein [Escherichia coli]  | 66    | 47      | 702         |
| 278       | 2      | 2830       | 1784      | gi 1488662      | phosphatase-associated protein [Bacillus subtilis]   | 66    | 48      | 1047        |
| 278       | 3      | 3830       | 2952      | gi 303560       | ORF271 [Escherichia coli]  | 66    | 45      | 879         |
| 279       | 2      | 3894       | 2218      | gi 1185289      | 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1- carboxylate synthase [Bacillus subtilis]  | 66    | 48      | 1677        |
| 288       | 4      | 2535       | 2275      | gi 1256625      | putative [Bacillus subtilis]   | 66    | 42      | 261         |
| 292       | 2      | 1133       | 942       | gi 1511604      | M. jannaschii predicted coding region M1651 [Methanococcus jannaschii]   | 66    | 30      | 192         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession  | Match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|------------------|--|-------|---------|-------------|
| 294       | 1      | 1116       | 559       | gi 216314        | esterase [Bacillus stearothermophilus]   | 66    | 45      | 558         |
| 297       | 4      | 2913       | 1978      | gi 994794        | cytochrome a assembly facto [Bacillus subtilis] sp P24009 COXK_BACSU PROBABLE CYTOCHROME C OXIDASE ASSEMBLY FACTOR.  | 66    | 45      | 936         |
| 316       | 4      | 2053       | 2682      | gi 1107839       | lactinate lyase [Pseudomonas aeruginosa]   | 66    | 40      | 630         |
| 338       | 4      | 2460       | 2302      | gi 520750        | biotin synthetase [Bacillus sphaericus]  | 66    | 59      | 159         |
| 339       | 1      | 1214       | 735       | gi 467468        | 7, 8-dihydro-6-hydroxymethylpterin-pyrophosphokinase [Bacillus subtilis]   | 66    | 52      | 480         |
| 363       | 1      | 3          | 863       | gi 581649        | lepiC gene product [Staphylococcus epidermidis]  | 66    | 47      | 863         |
| 366       | 2      | 232        | 483       | gi 1103505       | unknown [Schizosaccharomyces pombe]  | 66    | 53      | 252         |
| 367       | 4      | 2468       | 1845      | sp P20692 TYRA_  | PREPHENATE DEHYDROGENASE (EC 1.3.1.12) (PDH).  | 66    | 50      | 624         |
| 372       | 3      | 2150       | 1599      | gi 467416        | unknown [Bacillus subtilis]  | 66    | 38      | 552         |
| 378       | 1      | 212        | 1009      | gi 147309        | purine nucleoside phosphorylase [Escherichia coli]   | 66    | 50      | 798         |
| 401       | 1      | 1          | 462       | gi 388263        | p-aminobenzoic acid synthase [Streptomyces griseus] pir JN0531 JN0531 p-aminobenzoic acid synthase - Streptomyces riseus   | 66    | 46      | 462         |
| 404       | 7      | 4826       | 5254      | gi 606744        | cytidine deaminase [Bacillus subtilis]   | 66    | 51      | 429         |
| 411       | 2      | 1738       | 1103      | gi 1460081       | unknown [Mycobacterium tuberculosis]   | 66    | 44      | 636         |
| 420       | 1      | 2          | 541       | gi 1046024       | Na+ ATPase subunit J [Mycoplasma genitalium]   | 66    | 49      | 540         |
| 431       | 1      | 1          | 858       | gi 1500008       | M. Jannaschii predicted coding region MJ1154 (Methanococcus Jannaschii)  | 66    | 50      | 858         |
| 443       | 7      | 5679       | 5299      | gi 852076        | MrgA [Bacillus subtilis]   | 66    | 46      | 381         |
| 464       | 3      | 3405       | 2413      | gi 153047        | lysostaphin (ttg start codon) [Staphylococcus simulans] pir A25883 A25883 lysostaphin precursor - Staphylococcus simulans sp P10547 LSTP_STASI LYOSTAPHIN PRECURSOR (EC 3.5.1.-) | 66    | 51      | 993         |
| 561       | 1      | 956        | 480       | gi 1204905       | DNA-3-methyladenine glycosidase 1 [Haemophilus influenzae]   | 66    | 45      | 477         |
| 562       | 3      | 1066       | 1383      | gi 1046082       | M. genitalium predicted coding region MG172 [Mycoplasma genitalium]  | 66    | 52      | 318         |
| 576       | 1      | 11         | 724       | gi 305014        | ORF_0234 [Escherichia coli]  | 66    | 43      | 714         |
| 577       | 3      | 1190       | 503       | gi 1001353       | hypothetical protein [Synecocystis sp.]  | 66    | 52      | 288         |
| 584       | 1      | 2          | 331       | sp P24204 YEB_A_ | HYPOTHETICAL 46.7 KD PROTEIN IN NSBB-RUBB INTERGENIC REGION (ORF1).  | 66    | 48      | 330         |
| 592       | 1      | 1410       | 706       | gi 928839        | ORF266; putative [Lactococcus lactis phase BK5-T]  | 66    | 51      | 705         |
| 601       | 1      | 1413       | 720       | gi 1488695       | novel antigen: orf-2 [Staphylococcus aureus]   | 66    | 55      | 714         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 619       | 3      | 468        | 845       | gi 746573       | similar to H. musculus transport system membrane protein, N-ramp PIR:A40739 and S. cerevisiae SMP1 protein (PIR:A45154) Caenorhabditis elegans | 66    | 45      | 378         |
| 706       | 2      | 561        | 355       | gi 804808       | unknown protein (Rattus norvegicus)  | 66    | 46      | 207         |
| 734       | 2      | 673        | 512       | gi 1519085      | phosphatidylcholine binding immunoglobulin heavy chain IgM variable region (Mus musculus)  | 66    | 60      | 162         |
| 740       | 1      | 3          | 317       | gi 1209272      | argininosuccinate lyase (Campylobacter jejuni)   | 66    | 47      | 315         |
| 764       | 1      | 310        | 747       | gi 435296       | alkaline phosphatase like protein (Lactococcus lactis) pir S39339 S39339 alkaline phosphatase-like protein - Lactococcus actis                 | 66    | 42      | 418         |
| 852       | 1      | 338        | 171       | gi 536955       | CC Site No. 361 (Escherichia coli)   | 66    | 43      | 168         |
| 886       | 1      | 3          | 158       | gi 28272        | ferrichrome-binding protein (Bacillus subtilis)  | 66    | 41      | 156         |
| 889       | 1      | 462        | 232       | gi 1833061      | HCMVUL77 (AA 1-642) (Human cytomegalovirus)  | 66    | 66      | 231         |
| 893       | 1      | 2          | 247       | gi 149008       | putative (Halobacter pylori)   | 66    | 45      | 246         |
| 900       | 1      | 1425       | 733       | gi 580842       | P3 (Bacillus subtilis)   | 66    | 51      | 693         |
| 906       | 2      | 2300       | 1473      | gi 790945       | aryl-alcohol dehydrogenase (Bacillus subtilis)   | 66    | 53      | 828         |
| 947       | 1      | 79         | 549       | gi 410117       | diaminopimelate decarboxylase (Bacillus subtilis)  | 66    | 47      | 471         |
| 950       | 1      | 1100       | 552       | gi 48713        | orf145 (Staphylococcus aureus)   | 66    | 35      | 549         |
| 955       | 2      | 89         | 475       | gi 1204390      | uridine kinase (uridine monophosphokinase) (Haemophilus influenzae)  | 66    | 50      | 387         |
| 981       | 2      | 1108       | 997       | gi 457146       | rhoptry protein (Plasmodium yoelii)  | 66    | 18      | 312         |
| 986       | 1      | 25         | 315       | gi 305002       | ORF_4356 (Escherichia coli)  | 66    | 31      | 291         |
| 1057      | 1      | 3          | 203       | gi 1303853      | VQBF (Bacillus subtilis)   | 66    | 40      | 201         |
| 1087      | 1      | 1          | 294       | gi 575913       | unknown (Saccharomyces cerevisiae)   | 66    | 53      | 294         |
| 1105      | 1      | 1          | 231       | gi 1045799      | methylgalactoside permease ATP-binding protein (Mycoplasma genitalium)   | 66    | 46      | 231         |
| 1128      | 1      | 2          | 574       | gi 1001493      | hypothetical protein (Synchocystis sp.)  | 66    | 46      | 573         |
| 1150      | 1      | 498        | 250       | gi 1499034      | M. jannaschii predicted coding region M20255 (Methanococcus jannaschii)  | 66    | 40      | 249         |
| 1180      | 2      | 707        | 453       | gi 215908       | DNA polymerase (g43) (Bacteriophage T4)  | 66    | 46      | 255         |
| 1208      | 1      | 1123       | 587       | gi 1256653      | DNA-binding protein (Bacillus subtilis)  | 66    | 58      | 537         |
| 1342      | 1      | 1          | 402       | gi 1208474      | hypothetical protein (Synchocystis sp.)  | 66    | 53      | 402         |
| 1761      | 2      | 589        | 398       | gi 215811       | tall fiber protein (Bacteriophage T3)  | 66    | 50      | 192         |

TABLE 2

S aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 1983      | 1      | 499        | 251       | gi 1045935      | DNA helicase II (Mycoplasma genitalium)  | 66    | 40      | 249         |
| 2103      | 2      | 176        | 400       | gi 929798       | precursor for the major merozoite surface antigens (Plasmodium alciapum)   | 66    | 46      | 225         |
| 2341      | 1      | 373        | 188       | gi 1236623      | exodeoxyribonuclease (Bacillus subtilis)   | 66    | 38      | 186         |
| 2458      | 1      | 335        | 164       | gi 1019410      | unknown (Schizosaccharomyces pombe)  | 66    | 47      | 162         |
| 2505      | 1      | 468        | 235       | gi 1510394      | putative transcriptional regulator (Methanococcus jannaschii)  | 66    | 39      | 234         |
| 2525      | 1      | 558        | 280       | gi 1000695      | cytotoxin L (Clostridium sordellii)  | 66    | 44      | 279         |
| 2935      | 1      | 3          | 275       | gi 765073       | autolysin (Staphylococcus aureus)  | 66    | 47      | 273         |
| 3005      | 1      | 114        | 305       | gi 1205784      | heterocyst maturation protein (Haemophilus influenzae)   | 66    | 46      | 192         |
| 3048      | 1      | 80         | 277       | gi 1301813      | Yqew (Bacillus subtilis)   | 66    | 42      | 198         |
| 3071      | 1      | 1          | 189       | gi 1070014      | protein-dependent (Bacillus subtilis)  | 66    | 41      | 189         |
| 3081      | 1      | 404        | 225       | gi 984212       | unknown (Schizosaccharomyces pombe)  | 66    | 44      | 180         |
| 3090      | 2      | 580        | 386       | gi 1204987      | DNA polymerase III, alpha chain (Haemophilus influenzae)   | 66    | 48      | 195         |
| 3118      | 1      | 1          | 387       | gi 1009366      | respiratory nitrate reductase (Bacillus subtilis)  | 66    | 49      | 387         |
| 3739      | 1      | 786        | 400       | gi 1109684      | PreV (Bacillus subtilis)   | 66    | 47      | 399         |
| 3796      | 1      | 402        | 202       | gi 853760       | acyl-CoA dehydrogenase (Bacillus subtilis)   | 66    | 60      | 201         |
| 3934      | 1      | 59         | 347       | gi 563952       | glucanase perase (Bacillus licheniformis)  | 66    | 46      | 249         |
| 4240      | 1      | 3          | 350       | gi 151259       | HMC-CoA reductase (EC 1.1.1.88) (Pseudomonas mevalonii) p1r[A44756]A44756<br>hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.                     | 66    | 51      | 348         |
| 4604      | 1      | 7          | 234       | p1r[A26713]BHC  | hemocyanin subunit II - Atlantic horseshoe crab  | 66    | 46      | 228         |
| 4         | 9      | 8845       | 9750      | gi 145646       | cytR (Escherichia coli)  | 65    | 35      | 906         |
| 6         | 5      | 2708       | 3565      | gi 1887824      | ORE_0310 (Escherichia coli)  | 65    | 47      | 858         |
| 13        | 1      | 1993       | 998       | gi 143402       | recombination protein (ttr start codon) (Bacillus subtilis) gi 130392) RecN (Bacillus subtilis)  | 65    | 44      | 996         |
| 15        | 7      | 2493       | 3524      | gi 1403126      | cxd gene product (Alcaligenes eutrophus)   | 65    | 38      | 1032        |
| 18        | 3      | 1908       | 1372      | gi 149187       | acyltransferase (Saccharomyces cerevisiae)   | 65    | 50      | 537         |
| 21        | 3      | 1467       | 2482      | gi 149518       | phosphoribosyl anthranilate transferase (Lactococcus lactis) p1r[S5126]S5126 anthranilate phosphoribosyltransferase (EC 4.2.18) - Lactococcus lactis subsp. lactis | 65    | 52      | 1026        |
| 25        | 4      | 3374       | 4312      | gi 1502420      | malonyl-CoA:Acyl carrier protein transacylase (Bacillus subtilis)  | 65    | 44      | 939         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 27        | 2      | 390        | 626       | gi 1212729      | [Yqb3 (Bacillus subtilis)]   | 65    | 45      | 237         |
| 31        | 12     | 11040      | 11087     | gi 509245       | [D-hydroxylsuccinate dehydrogenase (Lactobacillus delbrueckii)]                        | 65    | 41      | 654         |
| 38        | 24     | 19172      | 19528     | gi 547          | [H-protein (Flavobacterium croquisat)]   | 65    | 41      | 357         |
| 44        | 2      | 790        | 1746      | gi 405882       | [yeK (Escherichia coli)]   | 65    | 46      | 957         |
| 44        | 12     | 9356       | 9832      | gi 1205905      | [molybdenum cofactor biosynthesis protein (Haemophilus influenzae)]                    | 65    | 50      | 525         |
| 45        | 8      | 6635       | 7588      | gi 493074       | [ApbA protein (Salmonella typhimurium)]  | 65    | 46      | 954         |
| 51        | 2      | 580        | 1503      | gi 580897       | [OppB gene product (Bacillus subtilis)]  | 65    | 45      | 924         |
| 52        | 1      | 225        | 953       | gi 1205518      | [NAD(P)H-flavin oxidoreductase (Haemophilus influenzae)]                               | 65    | 45      | 729         |
| 55        | 4      | 1139       | 1058      | pir A44459 A444 | [tropomyosin T beta TmT-5 - rabbit]  | 65    | 41      | 282         |
| 67        | 9      | 7421       | 8272      | gi 143607       | [apoptolysin protein (Bacillus subtilis)]  | 65    | 42      | 852         |
| 73        | 5      | 4466       | 5375      | gi 1204896      | [lysophospholipase L2 (Haemophilus influenzae)]  | 65    | 37      | 930         |
| 74        | 1      | 954        | 478       | gi 1204844      | [H. influenzae predicted coding region M10594 (Haemophilus influenzae)]                | 65    | 50      | 477         |
| 77        | 1      | 2          | 757       | gi 1046082      | [M. genitalium predicted coding region M0372 (Mycoplasma genitalium)]                  | 65    | 46      | 756         |
| 77        | 2      | 795        | 1433      | gi 1222116      | [permease (Haemophilus influenzae)]  | 65    | 37      | 639         |
| 81        | 3      | 4728       | 3454      | gi 1001708      | [hypothetical protein (Synecocystis sp.)]  | 65    | 49      | 1275        |
| 91        | 7      | 8548       | 8357      | gi 1399263      | [cystathionine beta-lyase (Emerella nidulans)]   | 65    | 40      | 192         |
| 98        | 3      | 1608       | 1988      | gi 467423       | [unknown (Bacillus subtilis)]  | 65    | 38      | 381         |
| 98        | 4      | 2250       | 2987      | gi 467424       | [unknown (Bacillus subtilis)]  | 65    | 45      | 738         |
| 102       | 3      | 2598       | 2119      | gi 1511532      | [N-terminal acetyltransferase complex, subunit ARD1 (Methanococcus jannaschii)]        | 65    | 39      | 480         |
| 102       | 4      | 3647       | 2862      | gi 1204637      | [H. influenzae predicted coding region M10388 (Haemophilus influenzae)]                | 65    | 32      | 786         |
| 103       | 9      | 10851      | 9841      | gi 142695       | [S-adenosyl-L-methionine:uroporphyrinogen III methyltransferase (Bacillus megaterium)] | 65    | 47      | 1011        |
| 103       | 10     | 10439      | 10119     | gi 710021       | [nitrite reductase (nirD) (Bacillus subtilis)]   | 65    | 51      | 321         |
| 106       | 2      | 262        | 1140      | gi 139881       | [ORP 311 (AA 1-311) (Bacillus subtilis)]   | 65    | 44      | 879         |
| 109       | 5      | 3909       | 4268      | gi 1204399      | [glucosamine-6-phosphate deaminase protein (Haemophilus influenzae)]                   | 65    | 44      | 360         |
| 109       | 10     | 7165       | 8595      | gi 536955       | [CG Site No. 361 (Escherichia coli)]   | 65    | 41      | 1431        |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 110       | 4      | 1688       | 3915      | gi 407881       | stringent response-like protein (Streptococcus equisimilis) p1r[S3975]S3975 stringent response-like protein - Streptococcus equisimilis   | 65    | 45      | 228         |
| 110       | 5      | 3882       | 4295      | gi 407880       | ORF1 (Streptococcus equisimilis)  | 65    | 50      | 414         |
| 110       | 6      | 4231       | 4380      | gi 1139574      | Orf2 (Streptomyces griseus)   | 65    | 56      | 150         |
| 112       | 110    | 9218       | 8640      | gi 1204571      | H. influenzae predicted coding region H10318 (Haemophilus influenzae)   | 65    | 52      | 579         |
| 112       | 112    | 12049      | 11288     | gi 1710496      | transcriptional activator protein (Bacillus brevis)   | 65    | 32      | 762         |
| 125       | 1      | 2          | 202       | gi 1151158      | repeat organellar protein (Plasmodium chabaudi)   | 65    | 39      | 201         |
| 126       | 1      | 3          | 422       | gi 137589       | precursor (Homo sapiens)  | 65    | 46      | 420         |
| 127       | 11     | 10711      | 12658     | gi 1064809      | homologous to sp:HTRA_ECOL1 (Bacillus subtilis)   | 65    | 41      | 1926        |
| 143       | 8      | 7543       | 7004      | gi 216513       | mutator mutT (AT-GC transversion) (Escherichia coli)  | 65    | 56      | 540         |
| 145       | 5      | 3587       | 3838      | gi 1209768      | D02_orf569 (Mycoplasma pneumoniae)  | 65    | 27      | 252         |
| 150       | 4      | 3482       | 2841      | gi 1146225      | putative (Bacillus subtilis)  | 65    | 37      | 642         |
| 166       | 1      | 3658       | 1948      | gi 148304       | beta-1,4-N-acetylmuramoylhydrolase (Enterococcus hirae) p1r[A42296]A42296 lysozyme 2 (EC 3.2.1.-) precursor - Enterococcus irae (ATCC 9790)   | 65    | 50      | 1911        |
| 188       | 6      | 3195       | 4178      | gi 151943       | ORF3 putative (Rhodobacter capsulatus)  | 65    | 46      | 984         |
| 189       | 9      | 4082       | 4785      | gi 58812        | ORF IV (AA 1-489) (F10001 mosaic virus)   | 65    | 40      | 108         |
| 195       | 6      | 7900       | 5272      | gi 145220       | alanyl-tRNA synthetase (Escherichia coli)   | 65    | 44      | 2637        |
| 195       | 7      | 10599      | 8104      | gi 482711       | exonuclease V alpha-subunit (Escherichia coli)  | 65    | 38      | 2496        |
| 206       | 16     | 16896      | 18191     | gi 408115       | ornithine acetyltransferase (Bacillus subtilis)   | 65    | 53      | 1296        |
| 217       | 4      | 3844       | 3215      | gi 1205974      | 5'guanylate kinase (Haemophilus influenzae)   | 65    | 41      | 630         |
| 220       | 4      | 5265       | 3751      | gi 580920       | rodd (grax) polypeptide (AA 1-673) (Bacillus subtilis) p1r[S06048]S06048 probable rodd protein - Bacillus subtilis sp:P3484[PAGE_BACSU PROBABLE POLY(GLYCEROL-PHOSPHATE) LPIA-GLUCOSYLTRANSFERASE (EC 2.4.1.52) (TECHIOIC ACID BIOSYNTHESIS ROUTIN E) | 65    | 40      | 1515        |
| 236       | 5      | 2327       | 3709      | gi 1146200      | DNA or RNA helicase, DNA-dependent ATPase (Bacillus subtilis)   | 65    | 46      | 1381        |
| 237       | 3      | 1902       | 2513      | gi 149379       | Hisad (Lactococcus lactis)  | 65    | 46      | 612         |
| 241       | 4      | 4968       | 4195      | gi 1205108      | ribonuclease H1 (EC 3.1.26.4) (RNASE H1) (Haemophilus influenzae)   | 65    | 50      | 774         |
| 252       | 1      | 1278       | 940       | gi 1204989      | hypothetical protein (GB:U00022.9) (Haemophilus influenzae)   | 65    | 40      | 339         |
| 261       | 5      | 4780       | 3794      | gi 145927       | fecD (Escherichia coli)   | 65    | 43      | 987         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 274       | 1      | 3          | 278       | gi1496558       | orfX [Bacillus subtilis]  | 65    | 42      | 276         |
| 301       | 2      | 982        | 815       | gi157118        | unknown [Bacillus subtilis]   | 65    | 45      | 168         |
| 307       | 4      | 3586       | 2864      | gi1070014       | protein-dependent [Bacillus subtilis]   | 65    | 40      | 723         |
| 335       | 2      | 2286       | 1399      | gi146913        | N-acetylglucosamine transport protein [Escherichia coli] pir109895 MOEC2N phosphotransferase system enzyme II (cc 7.1.69), N-acetylglucosamine-specific [Escherichia coli] sp109223 PMA_ECOLI PTS SYSTEM, N-ACETYLGLUCOSAMINE-SPECIFIC IIABC COMPONENT (EIIA) | 65    | 50      | 888         |
| 338       | 5      | 4120       | 3170      | gi1277029       | biotin synthase [Bacillus subtilis]   | 65    | 49      | 951         |
| 343       | 3      | 1490       | 2800      | gi143264        | membrane-associated protein [Bacillus subtilis]   | 65    | 48      | 1311        |
| 344       | 4      | 2161       | 2531      | gi1050540       | ERNA-glucanase synthetase [Lupinus luteus]  | 65    | 34      | 231         |
| 358       | 3      | 3421       | 3621      | gi1146220       | NAD <sup>+</sup> dependent glycerol-3-phosphate dehydrogenase [Bacillus subtilis]   | 65    | 47      | 201         |
| 364       | 1      | 238        | 699       | gi1340128       | ORF1 [Scaphylococcus aureus]  | 65    | 51      | 462         |
| 379       | 1      | 1          | 576       | gi143331        | alkaline phosphatase regulatory protein [Bacillus subtilis] pir127650 A27650 regulatory protein phor - Bacillus subtilis sp123545 PHOR_BACSU ALKALINE PHOSPHATASE SYNTHESIS SENSOR PROTEIN PHOR (EC 2.7.3.-)  | 65    | 40      | 576         |
| 379       | 3      | 3666       | 4346      | gi143268        | dihydroilpoamide transuccinylase (odhB; EC 2.3.1.61) [Bacillus subtilis]  | 65    | 50      | 681         |
| 428       | 1      | 187        | 483       | gi1420465       | ORF YOR195W [Saccharomyces cerevisiae]  | 65    | 45      | 297         |
| 434       | 2      | 272        | 810       | gi143498        | degS protein [Bacillus subtilis]  | 65    | 38      | 567         |
| 444       | 11     | 9280       | 10215     | gi1204756       | ribokinase [Haemophilus influenzae]   | 65    | 47      | 936         |
| 449       | 2      | 1241       | 1531      | gi1598848       | Na/H antiporter homolog [Lactococcus lactis]  | 65    | 41      | 291         |
| 478       | 2      | 1452       | 865       | gi1045942       | glycyl-tRNA synthetase [Hycoplasma genitalium]  | 65    | 39      | 588         |
| 479       | 1      | 1032       | 517       | gi1498192       | putative [Pseudomonas aeruginosa]   | 65    | 40      | 516         |
| 480       | 6      | 4312       | 5637      | gi1415662       | UDP-N-acetylglucosamine 1-carboxyvinyl transferase [Acinetobacter alcoaceticus]   | 65    | 48      | 1326        |
| 484       | 1      | 2          | 430       | gi1246551       | transmembrane protein (kdpD) [Escherichia coli]   | 65    | 44      | 429         |
| 499       | 1      | 54         | 932       | gi1603456       | reductase [Luishmania major]  | 65    | 53      | 879         |
| 505       | 1      | 914        | 459       | gi1518853       | OatA [Salmonella typhimurium]   | 65    | 39      | 456         |
| 571       | 2      | 1509       | 883       | gi149399        | open reading frame upstream glnE [Escherichia coli] ir157754 S37754 [hypochacal] protein XE (glnE 5' region) - cherichia coli   | 65    | 44      | 627         |
| 611       | 2      | 506        | 270       | gi10961         | RAP-2 [Plasmodium falciparum]   | 65    | 40      | 237         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 705       | 1      | 564        | 283       | gi 710020       | nitrite reductase (nirB) [Bacillus subtilis]   | 65    | 52      | 282         |
| 712       | 1      | 1          | 177       | gi 289272       | [ferrichrome-binding protein [Bacillus subtilis]   | 65    | 37      | 177         |
| 712       | 2      | 196        | 354       | gi 289272       | [ferrichrome-binding protein [Bacillus subtilis]   | 65    | 37      | 159         |
| 743       | 1      | 2          | 631       | gi 310631       | ATP binding protein [Streptococcus gordonii]   | 65    | 45      | 630         |
| 749       | 2      | 393        | 779       | gi 167374       | single strand DNA-binding protein [Bacillus subtilis]  | 65    | 29      | 387         |
| 762       | 1      | 1698       | 850       | gi 160399       | multidrug resistance protein [Pseudomonas fluorescens]   | 65    | 48      | 849         |
| 788       | 1      | 85         | 315       | gi 1129096      | unknown protein [Bacillus sp.]   | 65    | 35      | 231         |
| 850       | 1      | 1          | 408       | gi 1006604      | hypothetical protein [Synechocystis sp.]   | 65    | 37      | 408         |
| 908       | 1      | 1          | 444       | gi 1199546      | [2362] Saccharomyces cerevisiae  | 65    | 46      | 444         |
| 925       | 1      | 1          | 174       | gi 1256653      | DNA-binding protein [Bacillus subtilis]  | 65    | 54      | 174         |
| 1031      | 1      | 26         | 232       | gi 228657       | AppC-cytochrome d oxidase, subunit I homolog [Escherichia coli, K12, aptide, 514 aa]   | 65    | 47      | 207         |
| 1037      | 1      | 414        | 262       | gi 1491813      | gamma-glutamyltranspeptidase [Bacillus subtilis]   | 65    | 46      | 153         |
| 1053      | 1      | 348        | 175       | gi 642655       | unknown [Rhizobium meliloti]   | 65    | 34      | 174         |
| 1149      | 1      | 1399       | 752       | gi 1162980      | ribulose-5-phosphate 3-epimerase [Spinacia oleracea]   | 65    | 48      | 648         |
| 1214      | 1      | 881        | 495       | gi 1205959      | lactam utilization protein [Haemophilus influenzae]  | 65    | 45      | 387         |
| 1276      | 1      | 476        | 276       | pf 535493 5354  | site-specific DNA-methyltransferase SteI (EC 2.1.1.-) - Streptococcus sanguis  | 65    | 35      | 201         |
| 1276      | 2      | 900        | 577       | gi 471794       | [ORF] [Escherichia coli]   | 65    | 34      | 324         |
| 2057      | 1      | 272        | 118       | gi 633699       | TrnH [Yersinia enterocolitica]   | 65    | 21      | 135         |
| 2521      | 1      | 336        | 169       | gi 1045789      | hypothetical protein (CG:U4003.76) [Mycoplasma genitalium]   | 65    | 41      | 168         |
| 2974      | 1      | 590        | 297       | gi 152052       | enantiomerase-selective amidase [Rhodococcus sp.]  | 65    | 45      | 294         |
| 3031      | 1      | 306        | 154       | pf JQ1024 JQ10  | hypothetical JOK protein (JOK140 5' region) - fruit fly (Drosophila melanogaster)  | 65    | 45      | 153         |
| 3069      | 1      | 3          | 278       | gi 144906       | product homologous to E. coli thioredoxin reductase: J. Biol. Chem. 1988; 263:9015-9019, and to P52a protein of alkyl hydroperoxide oxidase from S. typhimurium: J. Biol. Chem. (1990) 265:10535-10540; pan reading frame A [Clostridium pasteurianum] | 65    | 46      | 276         |
| 3146      | 1      | 282        | 142       | gi 49315        | [ORF] gene product [Bacillus subtilis]   | 65    | 47      | 141         |
| 3170      | 1      | 679        | 341       | gi 1150711      | indolepyruvate decarboxylase [Erwinia herbicola]   | 65    | 44      | 339         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 3546      | 1      | 1          | 303       | gi1450688       | hsdM gene of Ecopri gene product [Escherichia coli] pir[S38437/S38437] hsdM protein - Escherichia coli pir[S09629/S09629] hypothetical protein A - Escherichia coli (SUB 40-520) | 65    | 42      | 303         |
| 3782      | 1      | 2          | 328       | gi1166412       | NADH-glutamate synthase (Medicago sativa)  | 65    | 42      | 327         |
| 3990      | 1      | 374        | 189       | gi11009166      | Respiratory nitrate reductase [Bacillus subtilis]  | 65    | 53      | 186         |
| 4032      | 1      | 613        | 308       | gi1123127       | ORF YGR087C [Saccharomyces cerevisiae]   | 65    | 50      | 306         |
| 4278      | 2      | 726        | 364       | gi11197667      | vitellogenin [Anolis pulchellus]   | 65    | 42      | 363         |
| 19        | 4      | 4259       | 5518      | gi1145727       | ideaD [Escherichia coli]   | 64    | 45      | 1260        |
| 19        | 6      | 7639       | 6926      | gi11016232      | lycF27 gene product [Cyanophora paradoxa]  | 64    | 36      | 714         |
| 20        | 8      | 7053       | 6454      | gi1765073       | autolysin [Staphylococcus aureus]  | 64    | 47      | 600         |
| 31        | 13     | 112706     | 11537     | gi1414009       | lipa-85d gene product [Bacillus subtilis]  | 64    | 45      | 1170        |
| 33        | 4      | 2388       | 4364      | gi11204696      | fructose-permease IIBC component [Haemophilus influenzae]  | 64    | 47      | 1972        |
| 36        | 3      | 1871       | 3013      | gi1290503       | glutamate permease [Escherichia coli]  | 64    | 40      | 1143        |
| 37        | 6      | 4065       | 4409      | gi139815        | orf2 gene product [Bacillus subtilis]  | 64    | 46      | 345         |
| 45        | 9      | 7852       | 8760      | gi11230585      | nucleotide sugar epimerase [Vibrio cholerae O139]  | 64    | 53      | 909         |
| 53        | 3      | 1540       | 1899      | gi11303961      | YQJ [Bacillus subtilis]  | 64    | 50      | 360         |
| 56        | 6      | 4793       | 3855      | gi1457514       | glcC [Bacillus subtilis]   | 64    | 45      | 919         |
| 56        | 124    | 30002      | 30247     | gi1470331       | similar to zinc finger [Caenorhabditis elegans]  | 64    | 42      | 246         |
| 62        | 4      | 2759       | 2421      | gi1642655       | unknown [Rhizobium meliloti]   | 64    | 28      | 339         |
| 85        | 6      | 7178       | 6027      | gi1457702       | 5-aminimidazole ribonucleotide-carboxylase [Pichia methanolica] pir[S39112/S39112] phosphoribosylaminoimidazole carboxylase (EC 1.1.21) - Yeast [Pichia methanolica]             | 64    | 46      | 1152        |
| 96        | 9      | 9251       | 10030     | gi1151513       | ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii]   | 64    | 42      | 780         |
| 100       | 1      | 1          | 600       | gi1765073       | autolysin [Staphylococcus aureus]  | 64    | 44      | 600         |
| 106       | 5      | 3868       | 4854      | gi1466778       | lysine specific permease [Escherichia coli]  | 64    | 46      | 987         |
| 123       | 2      | 838        | 554       | gi1467484       | unknown [Bacillus subtilis]  | 64    | 47      | 285         |
| 127       | 8      | 7514       | 7810      | gi1210061       | isotype-specific antigen [African horse sickness virus] pir[S27891/S27891] capsid protein vp2 - African horse sickness virus   | 64    | 28      | 297         |
| 131       | 7      | 7134       | 6721      | gi11511160      | M. jannaschii predicted coding region M1163 [Methanococcus jannaschii]   | 64    | 46      | 414         |

TABLE 2

S aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession   | match gene name  | V sim | V ident | length (nt) |
|-----------|--------|------------|-----------|---|--|-------|---------|-------------|
| 142       | 5      | 5455       | 4817      | gi11173517  | riboflavin synthase alpha subunit [Actinobacillus pleuropneumoniae]    | 64    | 44      | 639         |
| 143       | 1      | 709        | 356       | gi1429501329  | probable reductase protein - Leishmania major                          | 64    | 52      | 354         |
| 149       | 10     | 3555       | 3295      | gi1198151   | major surface antigen HSG2 [Pneumocystis carinii]                      | 64    | 44      | 261         |
| 154       | 4      | 3134       | 2107      | gi1984587   | DlnP [Escherichia coli]  | 64    | 50      | 828         |
| 161       | 5      | 3855       | 4880      | gi1903304   | ORF72 [Bacillus subtilis]  | 64    | 37      | 1026        |
| 165       | 1      | 33         | 791       | gi1467483   | unknown [Bacillus subtilis]  | 64    | 38      | 759         |
| 175       | 6      | 6355       | 474       | gi13072398  | phad gene product [Rhizobium meliloti]                                 | 64    | 42      | 1512        |
| 188       | 3      | 2042       | 2500      | gi1001961   | MHC class II analog [Staphylococcus aureus]                            | 64    | 45      | 459         |
| 195       | 14     | 11667      | 11446     | gi1396380   | No definition line found [Escherichia coli]                            | 64    | 47      | 222         |
| 206       | 15     | 16429      | 16938     | gi1304134   | argC [Bacillus stearothermophilus]                                     | 64    | 49      | 510         |
| 215       | 1      | 560        | 282       | gi1142359   | ORF 6 [Azotobacter vinelandii]   | 64    | 39      | 279         |
| 243       | 7      | 7818       | 6928      | gi1414014   | lpa-90d gene product [Bacillus subtilis]                               | 64    | 49      | 891         |
| 258       | 2      | 1330       | 845       | gi1664754   | P17 [Listeria monocytogenes]   | 64    | 38      | 486         |
| 259       | 1      | 462        | 232       | gi11459663  | M. jennaschii predicted coding region M0837 [Methanococcus jennaschii] | 64    | 52      | 231         |
| 263       | 6      | 6565       | 5567      | gi1142828   | aspartate semialdehyde dehydrogenase [Bacillus subtilis]               | 64    | 48      | 999         |
|           |        |            |           | sp1004797/BIAS_BACSU ASPARTATE-SEMIALDEHYDE DEHYDROGENASE (BC_2.1.11) (ASA DEHYDROGENASE) |  |       |         |             |
| 271       | 1      | 3          | 1163      | gi1467091   | hflX; B2235_C2_202 [Mycobacterium leprae]                              | 64    | 44      | 1161        |
| 280       | 1      | 173        | 1450      | gi1303839   | YqfR [Bacillus subtilis]   | 64    | 43      | 1278        |
| 293       | 1      | 2532       | 1267      | gi1147345   | primosomal protein n' [Escherichia coli]                               | 64    | 45      | 1266        |
| 295       | 2      | 742        | 1488      | gi1459266   | potential membrane spanning protein [Staphylococcus hominis]           | 64    | 39      | 747         |
|           |        |            |           | plr[S42932/S42932 potential membrane spanning protein - taphylococcus hominis]            |  |       |         |             |
| 301       | 5      | 1625       | 1486      | gi1580835   | lysine decarboxylase [Bacillus subtilis]                               | 64    | 35      | 180         |
| 315       | 4      | 5064       | 3949      | gi1141396   | quinol oxidase [Bacillus subtilis]                                     | 64    | 45      | 1116        |
| 321       | 1      | 1264       | 635       | gi1710496   | transcriptional activator protein [Bacillus brevis]                    | 64    | 41      | 630         |
| 333       | 5      | 4520       | 4239      | gi1314295   | ORF2; putative 19 kDa protein [Listeria monocytogenes]                 | 64    | 43      | 282         |
| 342       | 1      | 1          | 549       | gi1142940   | lfeA [Bacillus subtilis]   | 64    | 38      | 549         |
| 353       | 3      | 2878       | 2324      | gi1537049   | ORF_0470 [Escherichia coli]  | 64    | 44      | 555         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match      | accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|------------|-----------|---|-------|---------|-------------|
| 379       | 2      | 827        | 3658      | gi 525295  | A328      | oxoglutarate dehydrogenase (liponate) (EC 1.3.4.21) - Bacillus subtilis           | 64    | 47      | 2832        |
| 404       | 6      | 4429       | 4839      | gi 36933   | A369      | diacylglycerol kinase homolog - Streptococcus mutans                              | 64    | 35      | 411         |
| 407       | 1      | 2020       | 1133      | gi 969026  |           | orfX (Bacillus subtilis)  | 64    | 41      | 888         |
| 435       | 1      | 1109       | 591       | gi 1146177 |           | phosphotransferase system glucose-specific enzyme II (Bacillus subtilis)          | 64    | 44      | 519         |
| 443       | 6      | 4082       | 4798      | gi 147309  |           | purine nucleoside phosphorylase [Escherichia coli]                                | 64    | 51      | 717         |
| 450       | 2      | 1035       | 1604      | gi 406376  |           | ORF_0162 [Escherichia coli]   | 64    | 38      | 570         |
| 470       | 5      | 1680       | 6107      | gi 1369948 |           | host interacting protein (Bacteriophage B1)                                       | 64    | 45      | 4828        |
| 486       | 4      | 1911       | 1471      | gi 1205582 |           | ispermidine/putrescine transport system permease protein [Haemophilus influenzae] | 64    | 35      | 441         |
| 497       | 1      | 2217       | 1159      | sp P36929  | FNU_E     | FNU PROTEIN   | 64    | 38      | 1059        |
| 501       | 1      | 3          | 410       | gi 142450  |           | ahrc protein (Bacillus subtilis)  | 64    | 38      | 408         |
| 514       | 1      | 3          | 290       | gi 120496  |           | H. influenzae predicted coding region H10238 (Haemophilus influenzae)             | 64    | 34      | 288         |
| 551       | 4      | 3162       | 3323      | gi 1204511 |           | bacterioferitin comigratory protein (Haemophilus influenzae)                      | 64    | 41      | 162         |
| 603       | 4      | 759        | 956       | gi 755823  |           | hADH dehydrogenase F [Streptococcus americana]                                    | 64    | 35      | 198         |
| 653       | 2      | 940        | 746       | gi 1213234 |           | dicarboxylic amino acids dipep permease [Saccharomyces cerevisiae]                | 64    | 41      | 195         |
| 660       | 3      | 3401       | 2257      | sp P46133  | VIMAIL    | HYPOTHETICAL PROTEIN IN OG1 5 REGION (FRAGMENT)                                   | 64    | 39      | 1145        |
| 695       | 1      | 11         | 502       | gi 1001383 |           | hypothetical protein [Synechocystis sp.]  | 64    | 41      | 493         |
| 702       | 1      | 3          | 752       | gi 142865  |           | DNA primase [Bacillus subtilis]   | 64    | 46      | 750         |
| 826       | 1      | 1          | 339       | gi 971336  |           | arginyl tRNA synthetase [Bacillus subtilis]                                       | 64    | 50      | 339         |
| 838       | 1      | 1831       | 917       | gi 1354775 |           | pfoS/R [Treponema pallidum]   | 64    | 41      | 915         |
| 864       | 3      | 675        | 944       | gi 359833  |           | cyclomaltodextrin glucanotransferase [Bacillus stearothermophilus] i 359835       | 64    | 47      | 270         |
| 887       | 1      | 3          | 677       | gi 153002  |           | enterotoxin type E precursor [Staphylococcus aureus] p1r A28179 A28179            | 64    | 46      | 675         |
|           |        |            |           |            |           | enterotoxin E precursor - Staphylococcus aureus sp P12993 ETXE_57AAU              |       |         |             |
|           |        |            |           |            |           | ENTEROTOXIN TYPE E PRECURSOR (SEE)  |       |         |             |
| 928       | 2      | 1172       | 963       | gi 311976  |           | fibrinogen-binding protein [Staphylococcus aureus] p1r S34270 S34270              | 64    | 41      | 210         |
|           |        |            |           |            |           | fibrinogen-binding protein - Staphylococcus aureus                                |       |         |             |
| 1049      | 2      | 800        | 606       | gi 1049115 |           | Pap60 [Bacillus subtilis]   | 64    | 42      | 195         |
| 1067      | 2      | 999        | 748       | gi 1151072 |           | hAdh precursor [Haemophilus ducreyi]  | 64    | 50      | 252         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 1120      | 1      | 50         | 202       | gi142439        | ATP-dependent nuclease [Bacillus subtilis]  | 64    | 30      | 153         |
| 1125      | 1      | 751        | 377       | gi158168        | leplB gene product [Staphylococcus epidermidis]   | 64    | 44      | 375         |
| 1688      | 1      | 402        | 214       | pit1A01365      | transforming protein K-raa - mouse  | 64    | 47      | 189         |
| 2472      | 1      | 2          | 358       | gi1487282       | Na <sup>+</sup> -ATPase subunit J [Enterococcus hirae]  | 64    | 36      | 357         |
| 2989      | 1      | 520        | 356       | gi1304134       | argC [Bacillus stearothermophilus]  | 64    | 50      | 165         |
| 3013      | 1      | 630        | 352       | gi1551699       | cytochrome oxidase subunit I [Bacillus firmus]  | 64    | 51      | 279         |
| 3034      | 1      | 546        | 274       | gi11204349      | hypothetical protein (GB:DB-D90212_3) [Haemophilus influenzae]  | 64    | 50      | 273         |
| 3157      | 1      | 613        | 308       | gi11009366      | respiratory nitrate reductase [Bacillus subtilis]   | 64    | 46      | 306         |
| 3303      | 1      | 90         | 362       | gi11107839      | arginate lyase [Pseudomonas aeruginosa]   | 64    | 43      | 273         |
| 3852      | 2      | 82         | 288       | gi1216746       | D-lactate dehydrogenase [Lactobacillus plantarum]   | 64    | 42      | 207         |
| 3868      | 1      | 1          | 312       | gi1149435       | putative [Lactococcus lactis]   | 64    | 48      | 312         |
| 3918      | 1      | 660        | 331       | gi15532         | acetyl-CoA acyltransferase [Yarrowia lipolytica]  | 64    | 46      | 330         |
| 4000      | 1      | 112        | 378       | gi194688        | unknown [Saccharomyces cerevisiae]  | 64    | 44      | 267         |
| 4009      | 1      | 81         | 368       | gi139372        | grsB gene product [Bacillus brevis]   | 64    | 41      | 288         |
| 4166      | 1      | 2          | 349       | gi149435        | putative [Lactococcus lactis]   | 64    | 46      | 348         |
| 4366      | 1      | 2          | 307       | gi1216267       | ORF2 [Bacillus megaterium]  | 64    | 43      | 306         |
| 4457      | 1      | 2          | 400       | gi1197667       | vitellogenin [Anolis pulchellus]  | 64    | 43      | 399         |
| 11        | 3      | 1539       | 2438      | gi1418228       | ORF C [Staphylococcus aureus]   | 63    | 32      | 900         |
| 24        | 7      | 5611       | 5423      | gi1369943       | AI gene product [Bacteriophage B1]  | 63    | 34      | 183         |
| 29        | 1      | 1          | 390       | gi1467441       | expressed at the end of exponential growth under conditions in which the enzymes of the TCA cycle are repressed [Bacillus subtilis] gi1467441 | 63    | 43      | 390         |
| 31        | 6      | 6329       | 5712      | gi1496943       | ORF [Saccharomyces cerevisiae]  | 63    | 47      | 618         |
| 44        | 23     | 14669      | 15019     | pir1A04446      | hypothetical protein P-92 - Escherichia coli  | 63    | 36      | 351         |
| 48        | 6      | 4403       | 6250      | gi143498        | pyruvate synthase [Halobacterium halobium]  | 63    | 42      | 1848        |
| 50        | 5      | 3869       | 4738      | gi1413967       | lipA-43d gene product [Bacillus subtilis]   | 63    | 43      | 870         |
| 53        | 6      | 6764       | 5742      | gi1474176       | regulator protein [Staphylococcus xylosum]  | 63    | 49      | 1023        |

TABLE 2

3. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name  | % ident | Length (nt) |
|-----------|--------|------------|-----------|-----------------|--|---------|-------------|
| 56        | 14     | 15880      | 17607     | gi 467409       | DNA polymerase III subunit [Bacillus subtilis]   | 63      | 1728        |
| 57        | 11     | 7945       | 7376      | gi 37036        | ORF_0158 [Escherichia coli]  | 63      | 570         |
| 62        | 3      | 2479       | 2114      | gi 42656        | unknown [Rhizobium meliloti]   | 63      | 366         |
| 70        | 8      | 6562       | 7353      | gi 139821       | PhcC [Rhizobium meliloti]  | 63      | 792         |
| 75        | 2      | 223        | 927       | gi 149176       | HisG [Lactococcus lactis]  | 63      | 705         |
| 78        | 5      | 4912       | 4403      | gi 413950       | lpa-26d gene product [Bacillus subtilis]   | 63      | 510         |
| 91        | 5      | 9076       | 7220      | gi 466997       | metH2; B2126:C1_157 [Mycobacterium leprae]   | 63      | 1857        |
| 91        | 8      | 10566      | 9448      | gi 1206344      | cystathionine gamma-synthase [Haemophilus influenzae]  | 63      | 1119        |
| 120       | 1      | 21         | 1508      | gi 882657       | sulfite reductase (NADPH) [flavoprotein beta subunit [Escherichia coli]]   | 63      | 1488        |
| 120       | 4      | 2722       | 4125      | gi 665994       | hypothetical protein [Bacillus subtilis]   | 63      | 1404        |
| 127       | 7      | 6064       | 7566      | gi 40162        | murE gene product [Bacillus subtilis]  | 63      | 1503        |
| 149       | 6      | 2321       | 2106      | gi 148503       | dnaK [Escherichia coli]  | 63      | 216         |
| 149       | 26     | 10445      | 10170     | gi 4870         | ORF 2, has similarity to DNA polymerase (Saccharomyces kluyveri) r1515961/1515961 hypothetical protein 2 - yeast (Saccharomyces uverii) plasmid pSKL | 63      | 276         |
| 164       | 2      | 507        | 1298      | gi 145476       | CDP-diglyceride synthetase [Escherichia coli]  | 63      | 792         |
| 166       | 6      | 7009       | 8164      | gi 151912       | fructose enzyme II [Rhodospirillum rubrum]   | 63      | 1746        |
| 169       | 4      | 1704       | 1886      | gi 152886       | elongation factor Ts (tsf) [Spirillum citrii]  | 63      | 183         |
| 188       | 5      | 3145       | 2951      | gi 1334547      | GIV COI 114 grp 1B protein [Podospira anserina]  | 63      | 195         |
| 195       | 13     | 11767      | 112804    | gi 606100       | ORF_0335 [Escherichia coli]  | 63      | 1038        |
| 201       | 2      | 607        | 2283      | gi 433534       | arginyl-tRNA synthetase [Corynebacterium glutamicum] p1r/A09936/1A09936 arginine-tRNA ligase (EC 6.1.1.15) - orynebacterium glutamicum               | 63      | 1677        |
| 206       | 14     | 15893      | 16489     | gi 580828       | N-acetyl-glucanase-gamma-semialdehyde dehydrogenase [Bacillus subtilis]  | 63      | 597         |
| 220       | 5      | 7769       | 5766      | gi 216334       | secA protein [Bacillus subtilis]   | 63      | 2004        |
| 221       | 1      | 74         | 907       | gi 677945       | AppA [Bacillus subtilis]   | 63      | 834         |
| 227       | 3      | 944        | 1708      | gi 1510558      | isobutyrate synthase [Methanococcus jannaschii]  | 63      | 765         |
| 261       | 2      | 804        | 1070      | gi 486511       | ORF_YKR051c [Saccharomyces cerevisiae]   | 63      | 267         |
| 269       | 2      | 3606       | 1960      | gi 148221       | DNA-dependent ATPase, DNA helicase [Escherichia coli] p1r/J50137/1A50137 recQ protein - Escherichia coli   | 63      | 1647        |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 278       | 8      | 7417       | 8176      | gi 99273        | cystathionine gamma-synthase [Mycobacterium leprae] sp P46807 HET9_MYCUE<br>CISTATHIONINE GAMMA-SYNTHASE (EC 4.2.99.9) O-SUCCINYLHOMOSERINE (THIOL)-<br>LYASE | 63    | 41      | 1242        |
| 287       | 2      | 738        | 1733      | gi 405133       | putative [Bacillus subtilis]  |       |         |             |
| 295       | 1      | 2          | 748       | gi 1239983      | hypothetical protein [Bacillus subtilis]  | 63    | 38      | 996         |
| 328       | 3      | 2148       | 3134      | gi 45302        | carrier protein (AA 1 - 437) [Pseudomonas aeruginosa] ir S11497 S11497<br>branched-chain amino acid tRNA:port protein brab - eudomonas aeruginosa             | 63    | 41      | 747         |
| 362       | 2      | 1626       | 1216      | sp P35136 SERA  | D-3-PHOSPHOGLYCERATE DEHYDROGENASE (EC 1.1.1.95) [PCDH]   | 63    | 36      | 987         |
| 404       | 1      | 326        | 1051      | gi 1303816      | Yqe2 [Bacillus subtilis]  | 63    | 38      | 411         |
| 405       | 3      | 2101       | 1715      | gi 1303914      | YqhV [Bacillus subtilis]  | 63    | 35      | 726         |
| 406       | 1      | 451        | 227       | gi 142152       | sulfate permease (gtg start codon) [Synecococcus PCC6301]<br>pir A30101 GRVCS7 sulfate transport protein - Synecococcus sp. PCC 7942                          | 63    | 42      | 387         |
| 415       | 2      | 1048       | 2718      | gi 1205402      | transport ATP-binding protein [Haemophilus influenzae]  | 63    | 43      | 225         |
| 426       | 4      | 3575       | 2679      | gi 393268       | 29-kilodalton protein [Streptococcus pneumoniae] sp P42362 P29K_STRPN 29 KD<br>MEMBRANE PROTEIN IN PSA 5'REGION ORF1  | 63    | 39      | 897         |
| 505       | 3      | 1347       | 2195      | gi 1418959      | orf6 [Lactobacillus sake]   | 63    | 40      | 849         |
| 507       | 1      | 2          | 574       | gi 546917       | comK [Bacillus subtilis, E26, Peptide, 192 aa]  | 63    | 35      | 573         |
| 562       | 2      | 146        | 1084      | gi 43985        | nifS-like gene [Lactobacillus delbrueckii]  | 63    | 45      | 939         |
| 675       | 1      | 427        | 215       | gi 1510994      | serine aminotransferase [Methanococcus jannaschii]  | 63    | 29      | 213         |
| 686       | 1      | 3          | 230       | gi 517356       | nitrate reductase (NADH) [Lotus japonicus]  | 63    | 52      | 228         |
| 701       | 1      | 3          | 392       | gi 881940       | NorQ protein [Paracoccus denitrificans]   | 63    | 41      | 390         |
| 720       | 1      | 2          | 400       | gi 47168        | open reading frame [Streptomyces lividans]  | 63    | 35      | 399         |
| 779       | 1      | 571        | 287       | gi 1261932      | unknown [Mycobacterium tuberculosis]  | 63    | 41      | 285         |
| 907       | 1      | 22         | 321       | gi 149445       | ORF1 [Lactococcus lactis]   | 63    | 27      | 300         |
| 972       | 1      | 784        | 399       | gi 1511235      | M. jannaschii predicted coding region MJ232 [Methanococcus jannaschii]  | 63    | 27      | 396         |
| 1085      | 1      | 1154       | 618       | gi 1204277      | hypothetical protein (CB-U00019.14) [Haemophilus influenzae]  | 63    | 38      | 537         |
| 1094      | 1      | 3          | 542       | gi 720943       | urea amidolyase [Bacillus subtilis]   | 63    | 39      | 540         |
| 1108      | 1      | 3          | 482       | pir S49892 S498 | regulation protein - Bacillus subtilis  | 63    | 44      | 480         |
| 1113      | 1      | 1231       | 617       | gi 493017       | endocarditis specific antigen [Enterococcus faecalis]   | 63    | 45      | 615         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession  | match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|------------------|--|-------|---------|-------------|
| 1300      | 1      | 3          | 695       | [sp P33940 YQJH  | HYPOTHETICAL 54.3 KO PROTEIN IN ECO-ALXB INTERGENIC REGION.  | 63    | 46      | 693         |
| 1325      | 1      | 1          | 204       | [gi 928989       | pt00 protein [Borrelia burgdorferi]  | 63    | 30      | 204         |
| 1814      | 1      | 3          | 245       | [gi 1303914      | Yqhy [Bacillus subtilis]   | 63    | 34      | 243         |
| 2021      | 1      | 498        | 250       | [pir C33496 C334 | hlaC homolog - Bacillus subtilis   | 63    | 46      | 249         |
| 2325      | 1      | 2          | 193       | [gi 436132       | product is similar to tnpA of transposon tn554 from Staphylococcus aureus [Clostridium butyricum]  | 63    | 40      | 192         |
| 2335      | 1      | 1          | 195       | [gi 1184298      | flagellar H5-ring protein [Borrelia burgdorferi]   | 63    | 47      | 195         |
| 2406      | 1      | 451        | 227       | [gi 1041785      | rhostry protein [Plasmodium yoelii]  | 63    | 33      | 225         |
| 2961      | 2      | 136        | 360       | [gi 312483       | [carbamoyl-phosphate synthase (glutamine-hydrolysing) [Bacillus albidolysicus]                     | 63    | 52      | 225         |
| 2965      | 1      | 1          | 402       | [gi 1407784      | orf-1; novel antigen [Staphylococcus aureus]   | 63    | 50      | 402         |
| 2987      | 1      | 583        | 293       | [gi 1224069      | amidase [Moraxella catarrhalis]  | 63    | 35      | 291         |
| 2994      | 1      | 266        | 135       | [gi 836666       | [phosphoribosylformimino-praic ketoisomerase [Rhodobacter phaeoideis]                              | 63    | 51      | 132         |
| 3043      | 1      | 440        | 252       | [gi 1480237      | [phenylacetaldehyde dehydrogenase [Escherichia coli]   | 63    | 40      | 189         |
| 3078      | 1      | 609        | 400       | [gi 1487982      | [intrinsic membrane protein [Mycoplasma hominis]   | 63    | 36      | 210         |
| 3139      | 1      | 2          | 217       | [gi 439126       | [glutamate synthase (NADPH) [Aspergillus brasiliensis] pir A49916 A49916                           | 63    | 47      | 216         |
| 3625      | 1      | 793        | 198       | [gi 623073       | [glutamate synthase (NADPH) (EC 1.4.1.13) - aspergillus brasiliensis                               | 63    | 48      | 196         |
| 3658      | 1      | 1          | 399       | [gi 1303697      | [ORF160; putative [Bacteriophage phi-11]   | 63    | 37      | 399         |
| 3659      | 1      | 3          | 395       | [gi 1256135      | [YrbF [Bacillus subtilis]  | 63    | 48      | 393         |
| 3783      | 1      | 720        | 361       | [gi 1256902      | [pyruvate decarboxylase isozyme 2 (Swiss Prot. accession number P16467) [Saccharomyces cerevisiae] | 63    | 34      | 360         |
| 3900      | 1      | 338        | 171       | [sp P10537 ARYB_ | [BETA-AMYLASE (EC 3.2.1.2) (1,4-ALPHA-D-GLUCAN MALTOHYDROLASE).                                    | 63    | 54      | 168         |
| 4309      | 1      | 3          | 176       | [pir A37967 A379 | [neural cell adhesion molecule Ng-CAM precursor - chicken  | 63    | 57      | 174         |
| 4367      | 1      | 1          | 195       | [gi 1121932      | [Par6p gene product [Pichia pastoris]  | 63    | 30      | 195         |
| 4432      | 1      | 1          | 312       | [gi 131259       | [HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas nevaloni] pir A44756 A44756                          | 63    | 51      | 312         |
| 4468      | 1      | 6          | 308       | [gi 1296464      | [hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.                                 | 63    | 36      | 303         |
| 33        | 3      | 1411       | 2400      | [gi 151675       | [ATPase [Lactococcus lactis]   | 62    | 44      | 990         |
| 36        | 9      | 5985       | 6218      | [gi 1490521      | [tagatose 6-P kinase (Streptococcus mutans)  | 62    | 51      | 234         |
|           |        |            |           |                  | [HMSH] [Homo sapiens]  |       |         |             |

TABLE 2

5. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 37        | 1      | 2          | 721       | gi 1107531      | ceup gene product [Campylobacter coli]  | 62    | 33      | 720         |
| 38        | 15     | 10912      | 11589     | gi 1222058      | H. influenzae predicted coding region HN1279 [Haemophilus influenzae]   | 62    | 38      | 678         |
| 38        | 25     | 19526      | 20129     | gi 695280       | 10872 [Alcaligenes eutrophus]   | 62    | 41      | 804         |
| 57        | 2      | 2523       | 1780      | gi 171234       | 1071 [Haemophilus influenzae]   | 62    | 55      | 744         |
| 57        | 9      | 6646       | 6350      | gi 508174       | EF18 domain of PTS-dependent Gat transport and phosphorylation Escherichia coli   | 62    | 35      | 297         |
| 58        | 1      | 2          | 559       | gi 755152       | highly hydrophobic integral membrane protein [Bacillus subtilis] sp P42953 TAGC_BACSU TEICHOIC ACID TRANSLOCATION PERMEASE PROTEIN AGC.   | 62    | 34      | 558         |
| 67        | 10     | 8250       | 9014      | gi 470683       | Shows similarity with ATP-binding proteins from other ABC-transport perons, Swiss Prot Accession Numbers P24137, t08007, P04285, P24136 Escherichia coli                        | 63    | 31      | 765         |
| 69        | 8      | 8115       | 7494      | gi 46816        | actVA 4 gene product [Streptomyces coelicolor]  | 62    | 44      | 822         |
| 80        | 3      | 1793       | 1320      | gi 39993        | UDP-N-acetylmuramoylalanine-D-glutamate ligase [Bacillus subtilis]  | 62    | 43      | 474         |
| 87        | 7      | 7034       | 9205      | gi 217191       | 5'-nucleotidase precursor [Vibrio parahaemolyticus]   | 62    | 38      | 2172        |
| 100       | 3      | 4051       | 3089      | gi 1511047      | phosphoglycerate dehydrogenase [Methanococcus jannaschii]   | 62    | 42      | 963         |
| 102       | 1      | 2          | 520       | gi 153655       | mismatch repair protein [Streptococcus pneumoniae] pif[C28667 C28667 DNA mismatch repair protein hexa - Streptococcus pneumoniae]   | 62    | 34      | 519         |
| 112       | 2      | 466        | 1068      | gi 153741       | ATP-binding protein [Streptococcus mutans]  | 62    | 37      | 601         |
| 114       | 7      | 6855       | 7562      | gi 1204866      | L-fucose operon activator [Haemophilus influenzae]  | 62    | 38      | 708         |
| 116       | 4      | 6823       | 5633      | gi 677947       | AppC [Bacillus subtilis]  | 62    | 37      | 1191        |
| 124       | 8      | 6855       | 6004      | gi 853777       | product similar to E.coli PAF2 protein [Bacillus subtilis] pif[S55438 S55438 ywke protein - Bacillus subtilis sp P45873 HUNK_BACSU POSSIBLE PROTOPHYRINOGEN OXIDASE (EC 3.3.-1) | 62    | 44      | 852         |
| 148       | 1      | 24         | 554       | gi 467456       | unknown [Bacillus subtilis]   | 62    | 30      | 531         |
| 149       | 20     | 7591       | 6725      | gi 1205807      | replicative DNA helicase [Haemophilus influenzae]   | 62    | 41      | 867         |
| 163       | 3      | 1703       | 1153      | gi 40067        | X gene product [Bacillus sphaericus]  | 62    | 42      | 351         |
| 164       | 15     | 14673      | 15632     | gi 42219        | P15 gene product (AA 1 - 314) [Escherichia coli]  | 62    | 38      | 960         |
| 165       | 2      | 1166       | 1447      | gi 403936       | phenylalanyl-tRNA synthetase alpha subunit (Gly294 variant) unidentified cloning vector   | 62    | 38      | 282         |
| 166       | 2      | 2084       | 5089      | gi 308861       | CTC start codon [Lactococcus lactis]  | 62    | 44      | 3006        |
| 171       | 1      | 1225       | 614       | gi 1046053      | hypothetical protein (SP:P32049) [Mycoplasma genitalium]  | 62    | 41      | 612         |

TABLE 2

B. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 183       | 1      | 2521       | 1310      | gi 143045       | hemY [Bacillus subtilis]   | 62    | 45      | 1212        |
| 200       | 1      | 3          | 956       | gi 142439       | ATP-dependent nuclease [Bacillus subtilis]   | 62    | 32      | 954         |
| 237       | 2      | 935        | 1966      | gi 41695        | hlaC protein [Escherichia coli]  | 62    | 44      | 1032        |
| 261       | 3      | 4008       | 2605      | gi 143121       | ORF A1 putative [Bacillus firmus]  | 62    | 42      | 1404        |
| 289       | 8      | 4477       | 4719      | gi 467441       | expressed at the end of exponential growth under conditions in which he enzymes of the TCA cycle are repressed [Bacillus subtilis] gi 467441 expressed at the end of exponential growth under conditions in which the enzymes of the TCA cycle are repressed Bacil | 62    | 47      | 243         |
| 304       | 6      | 5018       | 3819      | gi 153015       | FenA protein [Staphylococcus aureus]   | 62    | 43      | 1200        |
| 324       | 1      | 2          | 262       | gi 142717       | cytochrome aa3 controlling protein [Bacillus subtilis] pir A33960 A33960 cta protein - Bacillus subtilis sp P12946 CTAA_BACSU CYTOCHROME AA3 CONTROLLING PROTEIN.  | 62    | 30      | 261         |
| 325       | 2      | 269        | 1207      | gi 581088       | methionyl-tRNA formyltransferase [Escherichia coli]  | 62    | 39      | 939         |
| 332       | 6      | 4894       | 4631      | gi 1499960      | uridine 5'-monophosphate synthase [Methanococcus jannaschii]   | 62    | 36      | 264         |
| 355       | 1      | 2          | 370       | gi 145925       | fecB [Escherichia coli]  | 62    | 32      | 369         |
| 365       | 8      | 6628       | 6804      | gi 413943       | lipa-19d gene product [Bacillus subtilis]  | 62    | 54      | 177         |
| 369       | 2      | 2744       | 1626      | pir A33577 A435 | regulatory protein pfor - Clostridium perfringens  | 62    | 42      | 1119        |
| 370       | 1      | 34         | 264       | gi 40665        | beta-glucosidase [Clostridium thermocellum]  | 62    | 37      | 231         |
| 415       | 3      | 2709       | 3176      | gi 1205401      | transport ATP-binding protein [Haemophilus influenzae]   | 62    | 35      | 468         |
| 429       | 1      | 1578       | 790       | gi 1046024      | Na+ ATPase subunit J [Mycoplasma genitalium]   | 62    | 40      | 789         |
| 444       | 2      | 704        | 1369      | gi 581510       | modulation gene; integral membrane protein; homology to Rhizobium eguminosarum nodJ [Rhizobium loti]   | 62    | 37      | 666         |
| 477       | 2      | 751        | 1869      | pir A48440 A484 | ring-infected erythrocyte surface antigen 2, RESA-2 - Plasmodium falciparum  | 62    | 44      | 1119        |
| 485       | 1      | 241        | 1707      | gi 117934       | betaine aldehyd dehydrogenase [Beta vulgaris]  | 62    | 43      | 1467        |
| 487       | 3      | 1141       | 1311      | gi 149445       | ORF1 [Lactococcus lactis]  | 62    | 31      | 171         |
| 494       | 2      | 1134       | 1313      | gi 166835       | ribulose biphosphate carboxylase/oxygenase activase [Arabidopsis thaliana]   | 62    | 37      | 180         |
| 518       | 1      | 193        | 882       | gi 153491       | O-methyltransferase [Streptomyces glaucescens]   | 62    | 39      | 690         |
| 534       | 2      | 369        | 2522      | gi 1480429      | putative transcriptional regulator [Bacillus stearothermophilus]   | 62    | 35      | 2154        |
| 551       | 6      | 4371       | 4820      | gi 511113       | ferric uptake regulation protein [Campylobacter jejuni]  | 62    | 37      | 450         |
| 574       | 1      | 1          | 570       | gi 153000       | enterotoxin B [Staphylococcus aureus]  | 62    | 43      | 570         |

TABLE 2

S. aureus - putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 590       | 2      | 344        | 1171      | gi 40367        | [ORF] Clostridium acetobutylicum  | 62    | 37      | 838         |
| 655       | 1      | 396        | 830       | gi 147195       | [phnB protein] Escherichia coli   | 62    | 44      | 435         |
| 656       | 1      | 2          | 478       | gi 1205451      | [cell division inhibitor] Haemophilus influenzae  | 62    | 36      | 477         |
| 676       | 1      | 692        | 348       | gi 1511613      | [methyl coenzyme M reductase system, component A2] Methanococcus jannaschii                         | 62    | 36      | 345         |
| 687       | 1      | 493        | 248       | gi 49272        | [Asparaginase] Bacillus licheniformis   | 62    | 48      | 246         |
| 700       | 2      | 267        | 944       | gi 1205822      | [hypothetical protein (G8:X75627_4)] Haemophilus influenzae   | 62    | 40      | 678         |
| 840       | 2      | 1715       | 1041      | gi 1045865      | [M. genitalium predicted coding region NC18] Mycoplasma genitalium                                  | 62    | 36      | 675         |
| 864       | 4      | 898        | 1491      | gi 1144332      | [deoxyuridine nucleoside dehydrogenase] Homo sapiens  | 62    | 38      | 594         |
| 916       | 1      | 35         | 400       | gi 411931       | [lipa-7d gene product] Bacillus subtilis  | 62    | 45      | 366         |
| 1071      | 1      | 1          | 771       | gi 1510649      | [aspartokinase I] Methanococcus jannaschii  | 62    | 40      | 771         |
| 1084      | 1      | 19         | 609       | gi 688011       | [Agx-1 antigen] human, infantile patient, testis, Peptide, 505 aa                                   | 62    | 39      | 591         |
| 1103      | 1      | 3          | 203       | gi 581261       | [ORF homologous to g.coli mcrB] Herpetosiphon aurantiacus   | 62    | 51      | 201         |
| 1217      | 1      | 463        | 233       | gi 460025       | [ORF2, putative] Streptococcus pneumoniae   | 62    | 41      | 231         |
| 1533      | 1      | 644        | 414       | gi 411968       | [lipa-44d gene product] Bacillus subtilis   | 62    | 48      | 231         |
| 1537      | 1      | 3          | 257       | gi 1510641      | [alanine-RNA synthetase] Methanococcus jannaschii   | 62    | 29      | 255         |
| 2287      | 1      | 3          | 161       | gi 485956       | [imjC gene product] Proteus mirabilis   | 62    | 45      | 159         |
| 2386      | 1      | 3          | 245       | gi 285708       | [nontoxic component] Clostridium botulinum  | 62    | 31      | 243         |
| 2484      | 1      | 331        | 167       | gi 142092       | [DNA-repair protein (recA)] Anabaena variabilis   | 62    | 35      | 165         |
| 2490      | 1      | 798        | 400       | gi 581648       | [lepta gene product] Staphylococcus epidermidis   | 62    | 42      | 399         |
| 3016      | 1      | 596        | 300       | gi 710022       | [uroperiplinogen III] Bacillus subtilis   | 62    | 51      | 297         |
| 3116      | 1      | 1          | 213       | gi 466883       | [nifs: B1496_C2.193] Mycobacterium leprae   | 62    | 48      | 213         |
| 3297      | 1      | 823        | 413       | gi 475715       | [acetyl coenzyme A acetyltransferase (thiolase)] Clostridium acetobutylicum                         | 62    | 42      | 411         |
| 3609      | 1      | 31         | 276       | gi 1408301      | [homologous to N-acyl-L-amino acid amidohydrolase of Bacillus stearothermophilus] Bacillus subtilis | 62    | 48      | 246         |
| 3665      | 2      | 584        | 402       | gi 151259       | [HMG-CoA reductase (EC 1.1.1.88)] Pseudomonas nevalonii   | 62    | 40      | 183         |
| 3733      | 1      | 3          | 374       | gi 1351197      | [thioredoxin reductase] Eubacterium acidaminophilum   | 62    | 42      | 372         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 3898      | 1      | 1          | 237       | gi153675        | tagatase 6-P kinase [Streptococcus mutans]   | 62    | 45      | 237         |
| 4027      | 1      | 283        | 143       | gi1310705       | homologue to gene 30 (aa 1-59); putative [bovine herpesvirus 4]  | 62    | 43      | 141         |
| 4109      | 1      | 727        | 365       | gi141748        | hsdN protein (AA 1-520) [Escherichia coli]   | 62    | 45      | 363         |
| 4303      | 1      | 1          | 303       | gi1303813       | YgeW [Bacillus subtilis]   | 62    | 43      | 303         |
| 4380      | 1      | 530        | 267       | gi1235684       | nevalonate pyrophosphate decarboxylase [Saccharomyces cerevisiae]  | 62    | 55      | 264         |
| 4494      | 1      | 2          | 256       | gi1510692       | enterotoxin H [Staphylococcus aureus]  | 62    | 34      | 255         |
| 4598      | 1      | 411        | 223       | gi1763513       | ORF4; putative [Streptomyces violaceoruber]  | 62    | 45      | 189         |
| 4624      | 1      | 1          | 222       | gi141748        | hsdN protein (AA 1-520) [Escherichia coli]   | 62    | 45      | 222         |
| 5         | 5      | 4288       | 3912      | gi1928831       | ORF95; putative [Lactococcus lactis phage BK5-7]   | 61    | 36      | 357         |
| 11        | 1      | 320        | 182       | pir[C33356]C333 | prothymosin alpha homolog (clone J2) - human (fragment)  | 61    | 33      | 159         |
| 16        | 11     | 10931      | 11938     | gi1205391       | [hypothetical] protein (SP:P3395) [Haemophilus influenzae]   | 61    | 44      | 948         |
| 32        | 1      | 281        | 401       | gi1066504       | exo-beta 1,3 glucanase [Cochliobolus carbonum]   | 61    | 50      | 519         |
| 38        | 3      | 616        | 1107      | gi1510864       | glutamine transport ATP-binding protein Q [Methanococcus jannaschii]   | 61    | 41      | 492         |
| 45        | 4      | 3082       | 4038      | gi1109686       | ProX [Bacillus subtilis]   | 61    | 45      | 957         |
| 48        | 8      | 7118       | 7504      | gi1498839       | ORF2 [Clostridium perfringens]   | 61    | 33      | 387         |
| 51        | 9      | 4605       | 5570      | gi1368269       | traC [Plasmid pAD1]  | 61    | 42      | 966         |
| 60        | 6      | 1689       | 2243      | gi1205893       | [hypothetical] protein (GB:U00011_31) [Haemophilus influenzae]   | 61    | 32      | 555         |
| 62        | 9      | 5559       | 5122      | gi1854656       | Na/H antiporter system ORF2 [Bacillus alcalophilus]  | 61    | 38      | 438         |
| 67        | 5      | 4330       | 5646      | gi1466612       | nika [Escherichia coli]  | 61    | 36      | 1317        |
| 74        | 2      | 2400       | 1504      | gi1204846       | carbamate kinase [Haemophilus influenzae]  | 61    | 40      | 897         |
| 85        | 3      | 2188       | 1101      | gi11498756      | amidophosphoribosyltransferase PurF [Rhizobium etli]   | 61    | 41      | 1098        |
| 86        | 4      | 1995       | 1582      | gi11499931      | M. jannaschii predicted coding region M1083 [Methanococcus jannaschii]   | 61    | 44      | 414         |
| 97        | 3      | 74         | 649       | gi1518679       | orf [Bacillus subtilis]  | 61    | 44      | 576         |
| 99        | 2      | 2454       | 1990      | gi1413958       | [ipa-34d gene product (Bacillus subtilis)]   | 61    | 18      | 465         |
| 124       | 7      | 6223       | 5123      | gi1556881       | [similar to Saccharomyces cerevisiae SNA5 protein (Bacillus subtilis)]<br>pir[S4938]S4938 ipc-29d protein - Bacillus subtilis sp[93953]YMLC_BACSU<br>HYPOTHETICAL 37.0 KD PROTEIN IN SP0118-GLYC NTERGENIC REGION. | 61    | 46      | 1101        |
| 125       | 4      | 1668       | 2531      | gi11491643      | ORFA gene product [Chloroflexus aurantiacus]   | 61    | 43      | 864         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 132       | 1      | 1250       | 627       | gi 1402559 P002 | hypothetical protein 1 - Enterococcus faecalis plasmid pAM-beta-1 (fragment)  | 61    | 43      | 624         |
| 149       | 9      | 3617       | 3075      | gi 1144332      | deoxyuridine nucleotidohydrolase (Homo sapiens)   | 61    | 40      | 543         |
| 149       | 12     | 8690       | 7869      | gi 160047       | p101/acidic basic repeat antigen (Plasmodium falciparum) p101/acidic basic repeat antigen precursor - Plasmodium falciparum (strain Camp)   | 61    | 35      | 822         |
| 168       | 1      | 1915       | 2361      | gi 1499694      | HIT protein, member of the HIT-family (Methanococcus jannaschii)  | 61    | 41      | 447         |
| 171       | 9      | 9675       | 7948      | gi 1467436      | similar to SpoVB (Bacillus subtilis)  | 61    | 38      | 1728        |
| 174       | 3      | 1042       | 2140      | gi 216374       | glucaryl 7-ACH acylase precursor (Bacillus laterosporus)  | 61    | 49      | 1299        |
| 190       | 4      | 5034       | 4111      | gi 409286       | bmrU (Bacillus subtilis)  | 61    | 37      | 924         |
| 216       | 1      | 2          | 190       | gi 415861       | eukaryotic initiation factor 2 beta (eIF-2 beta) (Cryptosporidium parvum)   | 61    | 29      | 189         |
| 227       | 7      | 4181       | 5048      | gi 216341       | ORF for methionine amino peptidase (Bacillus subtilis)  | 61    | 41      | 888         |
| 238       | 4      | 1959       | 3047      | gi 1409343      | CbrC protein (Erwinia chrysanthemi)   | 61    | 38      | 1089        |
| 247       | 1      | 2          | 694       | gi 537231       | ORF_579 (Escherichia coli)  | 61    | 38      | 693         |
| 247       | 2      | 678        | 1034      | gi 142226       | chvD protein (Agrobacterium tumefaciens)  | 61    | 40      | 357         |
| 257       | 2      | 3523       | 2627      | gi 699379       | glvr-1 protein (Mycobacterium leprae)   | 61    | 40      | 897         |
| 268       | 2      | 3419       | 3051      | gi 40364        | ORF1 (Clostridium acetobutylicum)   | 61    | 41      | 369         |
| 275       | 4      | 4621       | 4827      | gi 1204848      | hypothetical protein (GP:M07049.57) (Haemophilus influenzae)  | 61    | 36      | 207         |
| 277       | 1      | 1          | 1845      | gi 784897       | beta-N-acetylhexosaminidase (Streptococcus pneumoniae) p101/acidic basic repeat antigen precursor - Streptococcus pneumoniae  | 61    | 45      | 1845        |
| 278       | 9      | 8003       | 7032      | gi 467462       | cysteine synthetase A (Bacillus subtilis)   | 61    | 43      | 972         |
| 278       | 10     | 9878       | 8535      | gi 1205919      | Na <sup>+</sup> and Cl <sup>-</sup> dependent gamma-aminobutyric acid transporter (Haemophilus influenzae)  | 61    | 38      | 1344        |
| 283       | 1      | 1          | 366       | gi 755607       | polyA polymerase (Bacillus subtilis)  | 61    | 36      | 366         |
| 288       | 2      | 1918       | 1496      | gi 388108       | cell wall enzyme 1, enterococcus faecalis   | 61    | 43      | 423         |
| 291       | 1      | 86         | 334       | gi 454265       | PBP3 (Petunia hybrida)  | 61    | 38      | 249         |
| 318       | 1      | 1104       | 694       | gi 290531       | similar to beta-glucoside transport protein (Escherichia coli) sp P11451 PTIB_EC01_PTS_SYSTEM, ARBUTIN-LIKE IIB COMPONENT PHOSPHOTRANSFERASE ENZYME II, B COMPONENT (EC 2.7.1.69) | 61    | 47      | 411         |
| 330       | 2      | 1912       | 1190      | gi 1001805      | hypothetical protein (Synecocystis sp.)   | 61    | 41      | 723         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 385       | 2      | 1513       | 1025      | gi 533098       | DnaD protein [Bacillus subtilis]  | 61    | 42      | 489         |
| 426       | 1      | 794        | 399       | gi 1303853      | Yqef [Bacillus subtilis]  | 61    | 44      | 396         |
| 438       | 3      | 810        | 1421      | gi 1293660      | AbaA2 [Streptomyces coelicolor]   | 61    | 36      | 612         |
| 454       | 1      | 1580       | 792       | gi 733522       | phosphatidylinositol-4,5-diphosphate 3-kinase [Dictyostellum discoideum]  | 61    | 30      | 789         |
| 464       | 2      | 784        | 560       | gi 1123120      | IC53B7.5 gene product [Caenorhabditis elegans]  | 61    | 38      | 225         |
| 470       | 8      | 6077       | 7357      | gi 623073       | ORF360; putative [Bacteriophage LL-H]   | 61    | 47      | 1281        |
| 509       | 1      | 554        | 279       | gi 467484       | unknown [Bacillus subtilis]   | 61    | 45      | 276         |
| 555       | 3      | 1916       | 1296      | gi 141800       | anthranilate synthase glutamine amidotransferase [Acinetobacter alconecticus]   | 61    | 42      | 621         |
| 569       | 1      | 1711       | 857       | gi 467090       | R2235_C2_195 [Mycobacterium leprae]   | 61    | 47      | 855         |
| 585       | 2      | 961        | 803       | sp P36886 SUR2_ | SURVIVAL PROTEIN SUR2 HOMOLOG (FRAGMENT)  | 61    | 33      | 159         |
| 592       | 3      | 1694       | 1422      | gi 1221602      | immunity repressor protein [Haemophilus influenzae]   | 61    | 32      | 273         |
| 603       | 1      | 43         | 357       | gi 507738       | hmp [Vibrio parahaemolyticus]   | 61    | 33      | 315         |
| 669       | 1      | 2467       | 1235      | gi 1146243      | 22.4k identity with Escherichia coli DNA-damage inducible protein ...; putative [Bacillus subtilis]   | 61    | 37      | 1233        |
| 675       | 3      | 805        | 1101      | gi 403373       | glycerophosphoryl diester phosphodiesterase [Bacillus subtilis] pI[S37251]S37251 glycerophosphoryl diester phosphodiesterase - acillus subtilis | 61    | 36      | 297         |
| 703       | 1      | 1656       | 829       | gi 537181       | ORF_470 [Escherichia coli]  | 61    | 32      | 828         |
| 728       | 1      | 1628       | 816       | gi 806281       | DNA polymerase I [Bacillus stearothermophilus]  | 61    | 39      | 813         |
| 821       | 1      | 61         | 318       | gi 709992       | hypothetical protein [Bacillus subtilis]  | 61    | 38      | 258         |
| 856       | 2      | 2313       | 1567      | gi 609310       | portal protein gp3 [Bacteriophage HK97]   | 61    | 40      | 747         |
| 923       | 1      | 1081       | 542       | gi 143213       | putative [Bacillus subtilis]  | 61    | 38      | 540         |
| 1124      | 1      | 59         | 370       | gi 1107541      | C3D9.8 [Caenorhabditis elegans]   | 61    | 26      | 312         |
| 1492      | 1      | 548        | 276       | gi 406397       | unknown [Mycoplasma genitalium]   | 61    | 32      | 273         |
| 1602      | 1      | 46         | 318       | gi 733522       | phosphatidylinositol-4,5-diphosphate 3-kinase [Dictyostellum discoideum]  | 61    | 34      | 273         |
| 2500      | 1      | 577        | 290       | gi 1045964      | hypothetical protein (GB:U14001.297) [Mycoplasma genitalium]  | 61    | 31      | 288         |
| 2968      | 1      | 2          | 808       | gi 397526       | clumping factor [Staphylococcus aureus]   | 61    | 55      | 807         |
| 3076      | 1      | 3          | 248       | gi 143373       | ORF 1 [Lactococcus lactis]  | 61    | 41      | 246         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 3609      | 2      | 207        | 401       | gi 1408501      | homologous to N-acyl-L-amino acid amidohydrolase of <i>Bacillus stearothermophilus</i> [Bacillus subtilis]   | 61    | 39      | 195         |
| 3662      | 1      | 1477       | 740       | gi 1303813      | Yqew [Bacillus subtilis]   | 61    | 42      | 738         |
| 3672      | 1      | 2          | 442       | gi 784897       | beta-N-acetylhexosaminidase [Streptococcus pneumoniae] pir[A56390]A56390<br>mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96)<br>precursor - streptococcus pneumoniae | 61    | 50      | 441         |
| 3724      | 1      | 2          | 220       | gi 1009366      | Respiratory nitrate reductase [Bacillus subtilis]  | 61    | 41      | 219         |
| 3728      | 1      | 3          | 398       | gi 677843       | AppD [Bacillus subtilis]   | 61    | 46      | 396         |
| 3884      | 1      | 3          | 401       | gi 784897       | beta-N-acetylhexosaminidase [Streptococcus pneumoniae] pir[A56390]A56390<br>mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96)<br>precursor - streptococcus pneumoniae | 61    | 47      | 399         |
| 3971      | 1      | 3          | 383       | gi 784897       | beta-N-acetylhexosaminidase [Streptococcus pneumoniae] pir[A56390]A56390<br>mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96)<br>precursor - streptococcus pneumoniae | 61    | 45      | 381         |
| 4038      | 1      | 661        | 359       | gi 133970       | large subunit of NAM-dependent glucanase synthase [Plectonema boryanum]  | 61    | 24      | 303         |
| 4041      | 1      | 546        | 274       | gi 413953       | lpm-28d gene product [Bacillus subtilis]   | 61    | 48      | 273         |
| 4047      | 1      | 1          | 402       | gi 528991       | unknown [Bacillus subtilis]  | 61    | 42      | 402         |
| 4102      | 1      | 1          | 345       | gi 976025       | Hrsa [Escherichia coli]  | 61    | 46      | 345         |
| 4155      | 1      | 1          | 336       | gi 784897       | beta-N-acetylhexosaminidase [Streptococcus pneumoniae] pir[A56390]A56390<br>mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96)<br>precursor - streptococcus pneumoniae | 61    | 50      | 336         |
| 4268      | 1      | 463        | 233       | gi 450688       | hadH gene of Ecopri gene product [Escherichia coli] pir[S38437]S38437 hadH<br>protein - Escherichia coli pir[S09629]S09629 hypothetical protein A -<br>Escherichia coli (SUB 40-520)     | 61    | 38      | 231         |
| 4374      | 1      | 542        | 273       | gi 784897       | beta-N-acetylhexosaminidase [Streptococcus pneumoniae] pir[A56390]A56390<br>mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96)<br>precursor - streptococcus pneumoniae | 61    | 50      | 270         |
| 4389      | 1      | 2          | 172       | gi 147516       | ribokinase [Escherichia coli]  | 61    | 35      | 171         |
| 4621      | 1      | 2          | 268       | gi 784897       | beta-N-acetylhexosaminidase [Streptococcus pneumoniae] pir[A56390]A56390<br>mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96)<br>precursor - streptococcus pneumoniae | 61    | 47      | 267         |
| 4663      | 1      | 27         | 227       | gi 976025       | Hrsa [Escherichia coli]  | 61    | 50      | 201         |
| 4         | 6      | 6663       | 5536      | gi 1408501      | homologous to N-acyl-L-amino acid amidohydrolase of <i>Bacillus stearothermophilus</i> [Bacillus subtilis]   | 60    | 43      | 1128        |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | putative gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 11        | 6      | 3426       | 3725      | gi 410746       | ring-infected erythrocyte surface antigen (Plasmodium falciparum) pfrAD5526/AD5526 ring-infected erythrocyte surface antigen recursor - Plasmodium falciparum (strain FC27/Papua New Guinea) sp P13830 RESA_PLAUF RING-INFECTED ERYTHROCYTE SURFACE ANTIGEN RE | 60    | 24      | 300         |
| 11        | 14     | 11035      | 13011     | gi 1217651      | l-carbonyl reductase (NADPH) (Rattus norvegicus)   | 60    | 28      | 723         |
| 16        | 12     | 11917      | 12930     | gi 3001453      | hypothetical protein (Synchocystis sp.)  | 60    | 37      | 1014        |
| 33        | 1      | 26         | 469       | gi 386109       | regulatory protein (Enterococcus faecalis)   | 60    | 41      | 444         |
| 37        | 13     | 10414      | 9834      | gi 333656       | Ox1 (Bacillus subtilis)  | 60    | 40      | 981         |
| 39        | 4      | 4364       | 4522      | gi 4872         | ORF 4 (Saccharomyces kluyveri)   | 60    | 47      | 159         |
| 41        | 1      | 2047       | 1025      | gi 142822       | D-alanine racemase cds (Bacillus subtilis)   | 60    | 39      | 1033        |
| 43        | 4      | 2474       | 3607      | gi 468046       | para-nitrobenzyl esterase (Bacillus subtilis)  | 60    | 40      | 1134        |
| 44        | 10     | 6756       | 7769      | gi 414234       | lthIF (Escherichia coli)   | 60    | 52      | 1014        |
| 45        | 10     | 8874       | 9074      | gi 343949       | Ivar1(40.0) (Saccharomyces cerevisiae)   | 60    | 44      | 201         |
| 56        | 16     | 27842      | 26430     | gi 468764       | lmcR gene product (Rhizobium meliloti)   | 60    | 35      | 1413        |
| 60        | 2      | 173        | 388       | gi 3303864      | YggO (Bacillus subtilis)   | 60    | 33      | 216         |
| 63        | 2      | 357        | 1619      | gi 467124       | lureD, B229_C3.234 (Mycobacterium leprae)  | 60    | 43      | 1263        |
| 69        | 1      | 787        | 395       | gi 1518853      | OafA (Salmonella typhimurium)  | 60    | 36      | 193         |
| 88        | 1      | 1          | 1388      | gi 1480429      | putative transcriptional regulator (Bacillus stearothermophilus)   | 60    | 30      | 1188        |
| 92        | 6      | 4735       | 3881      | gi 349227       | transmembrane protein (Escherichia coli)   | 60    | 37      | 855         |
| 92        | 7      | 5996       | 4923      | gi 466613       | nlkB (Escherichia coli)  | 60    | 38      | 1074        |
| 93        | 1      | 949        | 476       | gi 1510925      | coenzyme F420-reducing hydrogenase, beta subunit (Methanococcus jannaschii)  | 60    | 27      | 474         |
| 96        | 6      | 7366       | 7478      | gi 472715       | acylase protein (Carnibacterium piscicola)   | 60    | 30      | 213         |
| 98        | 6      | 3212       | 4069      | gi 467425       | unknown (Bacillus subtilis)  | 60    | 42      | 858         |
| 102       | 10     | 7158       | 7430      | gi 141092       | acetolactate synthase small subunit (Bacillus subtilis) sp P37252 ILUN_BACSU ACETOLACTATE SYNTHASE SMALL SUBUNIT (EC 1.3.1.18) (AIAS) (ACETOHYDROXY-ACID SYNTHASE SMALL SUBUNIT) (ALS)   | 60    | 37      | 273         |
| 109       | 11     | 9127       | 10515     | gi 1255259      | D-succinylbenzoic acid (Dsb) CoA ligase (Staphylococcus aureus)  | 60    | 28      | 1389        |
| 109       | 12     | 10499      | 11656     | gi 141954       | beta-ketothiolase (Alcaligenes eutrophus)  | 60    | 41      | 1158        |
| 119       | 2      | 4630       | 3134      | gi 1524280      | unknown (Mycobacterium tuberculosis)   | 60    | 45      | 1497        |

TABLE 2

5. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 121       | 9      | 6957       | 7646      | gi1107529       | ceuC gene product (Campylobacter coli)  | 60    | 35      | 690         |
| 140       | 7      | 7704       | 6013      | gi146547        | kdpA (Escherichia coli)   | 60    | 45      | 1692        |
| 145       | 1      | 2          | 703       | gi11460077      | unknown (Mycobacterium tuberculosis)  | 60    | 23      | 702         |
| 150       | 3      | 2809       | 2216      | gi1146230       | putative (Bacillus subtilis)  | 60    | 40      | 594         |
| 157       | 2      | 1389       | 961       | gi11303975      | YojX (Bacillus subtilis)  | 60    | 30      | 429         |
| 158       | 5      | 5125       | 4769      | gi11449288      | unknown (Mycobacterium tuberculosis)  | 60    | 36      | 357         |
| 159       | 1      | 511        | 257       | gi1580932       | murD gene product (Bacillus subtilis)   | 60    | 43      | 255         |
| 160       | 1      | 159        | 1187      | gi11204532      | hypothetical protein (GB:U19201_21) (Haemophilus influenzae)  | 60    | 34      | 1029        |
| 161       | 14     | 4249       | 7866      | gi11496003      | ORF3: Putative oligonucleotide phosphatase based on homology with Lactococcus lactis Psp (Genbank Accession Number Z23527) (Caldicellulosiruptor saccharolyticus) | 60    | 34      | 384         |
| 172       | 3      | 1331       | 2110      | gi1485280       | 28.2 kDa protein (Streptococcus pneumoniae)   | 60    | 33      | 780         |
| 173       | 2      | 4082       | 2460      | gi11524197      | glycine betaine transporter OpuD (Bacillus subtilis)  | 60    | 41      | 1623        |
| 173       | 4      | 5963       | 4953      | gi1100737       | NAIP dependent leucokinin bc 12-hydroxydehydrogenase (Sus scrofa)   | 60    | 44      | 1011        |
| 198       | 3      | 3          | 995       | gi1413943       | lpa-19d gene product (Bacillus subtilis)  | 60    | 42      | 993         |
| 201       | 4      | 3641       | 4573      | sp P37028 YADT_ | HYPOTHETICAL 29.4 KD PROTEIN IN HPMU-PFS INTERGENIC REGION PRECURSOR.   | 60    | 37      | 933         |
| 203       | 3      | 3269       | 2415      | gi1927798       | [09719.34p: CAI: 0.14] (Saccharomyces cerevisiae)   | 60    | 43      | 855         |
| 206       | 9      | 12234      | 12515     | gi1937347 ECED_ | HYPOTHETICAL 21.4 KD PROTEIN IN ASP5 5'-REGION.   | 60    | 47      | 282         |
| 212       | 4      | 1213       | 1410      | gi1332711       | hemagglutinin-neuraminidase fusion protein (Human parainfluenza virus 3)  | 60    | 34      | 198         |
| 214       | 1      | 65         | 1153      | gi11204366      | hypothetical protein (GB:U14003_130) (Haemophilus influenzae)   | 60    | 36      | 1089        |
| 237       | 3      | 2          | 937       | gi1145377       | Msd (Lactococcus lactis)  | 60    | 40      | 936         |
| 241       | 6      | 5696       | 4998      | gi11046160      | hypothetical protein (GB:U00021_5) (Mycoplasma genitalium)  | 60    | 37      | 699         |
| 260       | 6      | 5919       | 6485      | gi1431950       | similar to a B-subtilis gene (GB: BACHEMENY_5) (Clostridium asteroidum)   | 60    | 35      | 567         |
| 264       | 1      | 2432       | 1218      | gi1397526       | clumping factor (Staphylococcus aureus)   | 60    | 53      | 1215        |
| 267       | 3      | 3          | 1409      | gi1148316       | NAH-antiporter protein (Enterococcus hirae)   | 60    | 27      | 1407        |
| 275       | 3      | 3804       | 4595      | gi1936889 F368  | leuD 3'-region hypothetical protein - Lactococcus lactis subsp. lactis [strain IL1403]  | 60    | 35      | 792         |
| 291       | 3      | 860        | 1198      | gi11208889      | coded for by C. elegans cDNA yk130e12.5; contains C2H2-type zinc fingers (Caenorhabditis elegans)   | 60    | 33      | 339         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 307       | 6      | 3421       | 3176      | gi 1070014      | protein-dependent [Bacillus subtilis]  | 60    | 36      | 246         |
| 316       | 8      | 4957       | 5823      | gi 413952       | ipa-28d gene product [Bacillus subtilis]   | 60    | 41      | 867         |
| 328       | 4      | 2996       | 3484      | gi 1204484      | membrane-associated component, branched amino acid transport system [Haemophilus influenzae]   | 60    | 39      | 489         |
| 332       | 5      | 4887       | 4363      | gi 1205449      | colicin V production protein (pur regulon) [Haemophilus influenzae]  | 60    | 37      | 525         |
| 357       | 1      | 1062       | 532       | gi 487842       | single-stranded DNA-specific exonuclease [Escherichia coli]  | 60    | 41      | 531         |
| 375       | 2      | 96         | 362       | gi 1057         | adenylyl cyclase gene product [Saccharomyces kluyveri] rJQ1145 OVBK<br>adenylyl cyclase [EC 4.6.1.1] - yeast [Saccharomyces kluyveri]  | 60    | 47      | 267         |
| 397       | 1      | 66         | 416       | gi 709999       | glucanate dehydratase [Bacillus subtilis]  | 60    | 37      | 351         |
| 409       | 1      | 2          | 163       | gi 499700       | glycogen phosphorylase [Saccharomyces cerevisiae]  | 60    | 35      | 162         |
| 453       | 4      | 914        | 1237      | gi 1196899      | unknown protein [Staphylococcus aureus]  | 60    | 36      | 324         |
| 453       | 7      | 3838       | 3620      | sp P12222 YCF1_ | HYPOHETICAL 226 KD PROTEIN (ORF 1901)  | 60    | 31      | 219         |
| 470       | 2      | 622        | 945       | pir S30782 S307 | Integrin homolog - yeast [Saccharomyces cerevisiae]  | 60    | 31      | 324         |
| 500       | 1      | 118        | 606       | gi 467407       | unknown [Bacillus subtilis]  | 60    | 36      | 469         |
| 503       | 3      | 752        | 982       | gi 167835       | myosin heavy chain [Dictyostelium discoideum]  | 60    | 34      | 231         |
| 505       | 4      | 2238       | 3563      | gi 1510732      | NADH oxidase [Methanococcus jannaschii]  | 60    | 26      | 1326        |
| 523       | 1      | 3          | 1043      | gi 143331       | alkaline phosphatase regulatory protein [Bacillus subtilis]<br>pir A27650 A27650 regulatory protein phor - Bacillus subtilis<br>sp P23545 PHOR_BACSU ALKALINE PHOSPHATASE SYNTHESIS SENSOR PROTEIN PHOR [EC 2.7.3.-] | 60    | 41      | 1041        |
| 543       | 1      | 1          | 465       | gi 1511103      | cobalt transport ATP-binding protein O [Methanococcus jannaschii]  | 60    | 40      | 465         |
| 545       | 1      | 1          | 726       | gi 1498192      | putative [Pseudomonas aeruginosa]  | 60    | 40      | 726         |
| 556       | 1      | 2          | 1054      | gi 1477402      | lex gene product [Bordetella pertussis]  | 60    | 42      | 1053        |
| 578       | 1      | 974        | 489       | gi 1205129      | H. influenzae predicted coding region H10882 [Haemophilus influenzae]  | 60    | 42      | 486         |
| 594       | 1      | 1          | 624       | gi 1212755      | adenylyl cyclase [Aeromonas hydrophila]  | 60    | 45      | 624         |
| 604       | 1      | 3          | 530       | gi 145925       | [fecB [Escherichia coli]   | 60    | 42      | 528         |
| 620       | 1      | 926        | 465       | gi 1205483      | bicyclic resistance protein [Haemophilus influenzae]   | 60    | 33      | 462         |
| 630       | 2      | 871        | 1122      | gi 1486242      | unknown [Bacillus subtilis]  | 60    | 41      | 252         |
| 645       | 2      | 574        | 425       | gi 1205136      | serine hydroxymethyltransferase [serine methylase] [Haemophilus influenzae]  | 60    | 28      | 150         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 684       | 1      | 1082       | 843       | gi1205538       | hypothetical protein (G0:U14003.102) [Haemophilus influenzae]   | 60    | 39      | 240         |
| 786       | 1      | 967        | 485       | gi1102944       | orfM1 gene product [Bacillus subtilis]  | 60    | 46      | 483         |
| 844       | 1      | 588        | 346       | gi1750943       | urea amidolyase [Bacillus subtilis]   | 60    | 40      | 243         |
| 851       | 1      | 1          | 726       | gi1159661       | GMP reductase [Ascaris lumbricoides]  | 60    | 41      | 726         |
| 871       | 1      | 1746       | 874       | gi11001493      | hypothetical protein [Synchocystis sp.]   | 60    | 39      | 873         |
| 896       | 1      | 1358       | 839       | gi1604926       | NADH dehydrogenase, subunit 5 [Schizosaccharomyces pombe] sp[P50366]NUSK_SCHCO<br>NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5 (EC 6.5.3)   | 60    | 39      | 720         |
| 908       | 2      | 448        | 753       | gi1662880       | novel hemolytic factor [Bacillus cereus]  | 60    | 31      | 306         |
| 979       | 1      | 2          | 595       | gi11429255      | putative: orf1 [Bacillus subtilis]  | 60    | 30      | 594         |
| 1078      | 1      | 669        | 502       | gi1581055       | inner membrane copper tolerance protein [Escherichia coli] gi1871029<br>disulphide isomerase like protein [Escherichia coli] pir1547295/S47295<br>inner membrane copper tolerance protein - scherichia coli | 60    | 40      | 168         |
| 1112      | 1      | 1150       | 620       | gi1407885       | ORF3 [Streptomyces griseus]   | 60    | 34      | 531         |
| 1135      | 1      | 484        | 275       | gi1171407       | Vps8p [Saccharomyces cerevisiae]  | 60    | 36      | 210         |
| 1146      | 1      | 17         | 562       | gi1239981       | hypothetical protein [Bacillus subtilis]  | 60    | 36      | 546         |
| 1291      | 1      | 716        | 360       | pir1557530/S575 | carboxyl esterase - Acinetobacter calcoaceticus   | 60    | 30      | 357         |
| 1332      | 1      | 336        | 169       | gi1222056       | aminotransferase [Haemophilus influenzae]   | 60    | 44      | 168         |
| 1429      | 1      | 3          | 146       | gi1205619       | ferritin like protein [Haemophilus influenzae]  | 60    | 39      | 144         |
| 1722      | 1      | 570        | 286       | gi1240052       | dlhydroflavonol-4-reductase, DFR [Hordeum vulgare-barley, cv. Gula, eptide, 354 aa]   | 60    | 36      | 285         |
| 2350      | 1      | 385        | 200       | gi1497626       | ORF 1 [Plasmid pAO1]  | 60    | 20      | 186         |
| 2936      | 1      | 519        | 310       | gi1508981       | prephenate dehydratase [Bacillus subtilis]  | 60    | 48      | 210         |
| 3027      | 1      | 568        | 302       | gi11146199      | putative [Bacillus subtilis]  | 60    | 37      | 267         |
| 3084      | 1      | 20         | 206       | gi11407784      | orf-1; novel antigen [Staphylococcus aureus]  | 60    | 51      | 189         |
| 3135      | 1      | 2          | 226       | gi11046097      | cytochrome-cytochrome protein [Mycobacterium genitalium]  | 60    | 34      | 225         |
| 3603      | 1      | 368        | 186       | gi1510108       | mitochondrial long-chain enoyl-CoA hydratase/3-hydroxycyl-CoA dehydrogenase<br>alpha-subunit [Rattus norvegicus]  | 60    | 42      | 183         |
| 3645      | 1      | 486        | 244       | gi1151259       | HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas merringtonii] pir154756/A4756<br>hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.   | 60    | 42      | 243         |
| 3747      | 1      | 3          | 146       | gi1474192       | lucC gene product [Escherichia coli]  | 60    | 36      | 144         |

TABLE 2

S aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 3912      | 1      | 3          | 335       | gi 488695       | novel antigen: orf-2 (Staphylococcus aureus)   | 60    | 44      | 333         |
| 4072      | 1      | 3          | 272       | gi 405879       | yeiH (Escherichia coli)  | 60    | 33      | 270         |
| 4134      | 1      | 510        | 332       | gi 80656        | chemoreceptor protein (Rhizobium leguminosarum bv. viciae) gi 780656   | 60    | 28      | 359         |
| 4207      | 2      | 677        | 402       | gi 602031       | similar to trimethylamine dehydrogenase (EC 5.99.7) - Mycoplasma capricolum (S0C3) (fragment)  | 60    | 41      | 276         |
| 4243      | 1      | 127        | 324       | gi 899317       | peptide synthetase module (Microcystis aeruginosa) pifS49111[S49111] Probable amino acid activating domain - Microcystis aeruginosa (fragment) (S0B 144-528) | 60    | 42      | 198         |
| 4310      | 1      | 624        | 313       | gi 308980       | phoB (Bacillus subtilis)   | 60    | 28      | 312         |
| 4345      | 1      | 343        | 173       | gi 510108       | mitochondrial long-chain enoyl-CoA hydratase/3-hydroxycy-CoA ehydrogenase alpha-subunit (Rattus norvegicus)  | 60    | 42      | 171         |
| 4382      | 1      | 498        | 280       | gi 47382        | acyl-CoA-dehydrogenase (Streptomyces purpurascens)   | 60    | 48      | 219         |
| 4474      | 1      | 53         | 223       | gi 510108       | mitochondrial long-chain enoyl-CoA hydratase/3-hydroxycy-CoA ehydrogenase alpha-subunit (Rattus norvegicus)  | 60    | 42      | 171         |
| 23        | 4      | 4518       | 3523      | gi 426846       | VipB protein (Salmonella typhi)  | 59    | 39      | 996         |
| 33        | 2      | 707        | 1483      | gi 548604[S486  | hypothetical protein - Mycoplasma capricolum (S0C3) (fragment)   | 59    | 33      | 777         |
| 33        | 5      | 4651       | 5853      | gi 6721         | P5982.3 (Caenorhabditis elegans)   | 59    | 33      | 1203        |
| 17        | 2      | 3228       | 3299      | gi 142833       | ORF2 (Bacillus subtilis)   | 59    | 37      | 910         |
| 38        | 21     | 16784      | 16593     | gi 912576       | BiP (Phaeodactylum tricornutum)  | 59    | 40      | 192         |
| 52        | 3      | 2648       | 2349      | gi 536972       | ORF_090a (Escherichia coli)  | 59    | 44      | 300         |
| 54        | 12     | 14181      | 13402     | gi 4803940      | transcription regulator (Bacillus subtilis)  | 59    | 37      | 780         |
| 57        | 3      | 4397       | 3339      | gi 508176       | Cat-1-P-DH, NAD dependent (Escherichia coli)   | 59    | 40      | 1059        |
| 66        | 1      | 986        | 495       | gi 1303901      | YqjH (Bacillus subtilis)   | 59    | 34      | 492         |
| 67        | 7      | 6552       | 7460      | gi 912461       | nlkC (Escherichia coli)  | 59    | 37      | 909         |
| 70        | 7      | 5383       | 6366      | gi 1399822      | PhoD precursor (Rhizobium meliloti)  | 59    | 46      | 984         |
| 78        | 1      | 1          | 1449      | gi 971345       | unknown, similar to E.coli cardiolipin synthase (Bacillus subtilis) sp P45860 WIE_BACSU HYPOTHETICAL 58.2 PROTEIN IN NARI-ACDA NTERGENIC REGION.             | 59    | 39      | 1449        |
| 82        | 10     | 14329      | 15534     | gi 490328       | LORF F (unidentified)  | 59    | 44      | 1206        |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 89        | 2      | 1602       | 958       | gi 642801       | unknown [Saccharomyces cerevisiae]  | 59    | 32      | 645         |
| 96        | 4      | 4940       | 5473      | gi 1333802      | protein of unknown function [Rhodobacter capsulatus]  | 59    | 33      | 534         |
| 98        | 1      | 2          | 820       | gi 667421       | similar to B. subtilis DnaM [Bacillus subtilis]   | 59    | 34      | 819         |
| 119       | 1      | 166        | 1557      | gi 143122       | ORF B: putative [Bacillus firmus]   | 59    | 36      | 1392        |
| 120       | 10     | 6214       | 6756      | gi 15354        | ORF 55.9 [Bacteriophage T4]   | 59    | 39      | 543         |
| 120       | 16     | 112476     | 11510     | gi 1086575      | beta [Phisobium mallocti]   | 59    | 44      | 1035        |
| 123       | 1      | 386        | 195       | gi 984737       | catalase [Campylobacter jejuni]   | 59    | 38      | 192         |
| 130       | 1      | 370        | 645       | gi 1256634      | 25.8% identity over 120 aa with the Synenococcus sp. NpeV protein; putative [Bacillus subtilis] | 59    | 31      | 276         |
| 131       | 4      | 5278       | 5712      | gi 1510655      | hypothetical protein (SP:P42297) [Methanococcus jannaschii]                                     | 59    | 39      | 435         |
| 164       | 1      | 3          | 509       | gi 1001342      | hypothetical protein [Synchocystis sp.]   | 59    | 41      | 507         |
| 164       | 4      | 1529       | 2821      | gi 1205165      | hypothetical protein (SP:P37764) [Haemophilus influenzae]                                       | 59    | 35      | 1291        |
| 164       | 19     | 19641      | 21376     | gi 1001381      | hypothetical protein [Synchocystis sp.]   | 59    | 34      | 1734        |
| 173       | 3      | 4727       | 3717      | gi 1164121      | leukin-induced protein [Vigna radiata]  | 59    | 50      | 1011        |
| 179       | 2      | 2218       | 1688      | gi 143036       | unidentified gene product [Bacillus subtilis]   | 59    | 30      | 531         |
| 195       | 12     | 12669      | 11503     | gi 762778       | Nifs gene product [Anabaena azollae]  | 59    | 41      | 1167        |
| 201       | 5      | 4702       | 5670      | gi 1510240      | hemolysin [Methanococcus jannaschii]  | 59    | 32      | 969         |
| 201       | 7      | 5719       | 6315      | gi 1511456      | M. jannaschii predicted coding region M31437 [Methanococcus jannaschii]                         | 59    | 34      | 597         |
| 209       | 1      | 102        | 461       | gi 1204666      | hypothetical protein (GN:X73124.53) [Haemophilus influenzae]                                    | 59    | 42      | 360         |
| 214       | 3      | 1050       | 2234      | gi 151531       | 2-nitropropane dioxygenase [Miliopsalis saturnus]   | 59    | 36      | 1185        |
| 214       | 5      | 3291       | 4135      | gi 1303709      | YrkJ [Bacillus subtilis]  | 59    | 32      | 843         |
| 217       | 2      | 3181       | 2167      | gi 1200489      | dtp (CG Site No. 18430) [Escherichia coli]  | 59    | 44      | 1215        |
| 217       | 5      | 3078       | 3785      | gi 149382       | hlsA [Lactococcus lactis]   | 59    | 38      | 708         |
| 251       | 2      | 376        | 960       | gi 1303791      | YrkJ [Bacillus subtilis]  | 59    | 34      | 585         |
| 286       | 1      | 1621       | 812       | gi 146551       | transmembrane protein (kdpI) [Escherichia coli]   | 59    | 31      | 810         |
| 316       | 5      | 4978       | 3860      | gi 405879       | yeiH [Escherichia coli]   | 59    | 32      | 1119        |
| 370       | 3      | 600        | 761       | gi 1303794      | YqeM [Bacillus subtilis]  | 59    | 35      | 162         |

TABLE 2

S. aureus - putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 382       | 1      | 1009       | 506       | gi 547513       | orf3 [Haemophilus influenzae]   | 59    | 34      | 504         |
| 391       | 3      | 1620       | 1273      | gi 152901       | ORF 3 [Spirochaeta aurantia]  | 59    | 37      | 348         |
| 406       | 3      | 2805       | 1705      | gi 709992       | hypothetical protein [Bacillus subtilis]  | 59    | 34      | 1101        |
| 426       | 5      | 3802       | 3245      | gi 1204610      | Iron(III) dictrate transport ATP-binding protein FECE [Haemophilus influenzae]  | 59    | 36      | 558         |
| 429       | 2      | 1513       | 1148      | gi 1064809      | homologous to spinRA_ECOLI [Bacillus subtilis]  | 59    | 42      | 366         |
| 460       | 2      | 708        | 1301      | gi 466882       | [ppsi: B1496_C2_185] [Mycobacterium leprae]   | 59    | 37      | 594         |
| 461       | 4      | 2112       | 3135      | gi 1498295      | homoserine kinase homolog [Streptococcus pneumoniae]  | 59    | 37      | 924         |
| 473       | 1      | 2929       | 1607      | gi 147989       | trigger factor [Escherichia coli]   | 59    | 40      | 1123        |
| 480       | 8      | 5862       | 6110      | gi 1205311      | [JRI]-hydroxymyristol acyl carrier protein dehydrase [Haemophilus influenzae]   | 59    | 40      | 249         |
| 521       | 1      | 14         | 1354      | gi A256201A256  | staphylococcalase - Staphylococcus aureus (fragment)  | 59    | 32      | 1341        |
| 534       | 4      | 2994       | 4073      | gi 153746       | mammotol-phosphate dehydrogenase [Streptococcus mutans] pif-[C44799][C44798] mannitol-phosphate dehydrogenase MtID - treptococcus mutans                    | 59    | 36      | 1080        |
| 535       | 1      | 1          | 954       | gi 1469939      | group B oligopeptidase Pepp [Streptococcus agalactiae]  | 59    | 33      | 954         |
| 551       | 3      | 2836       | 3186      | gi 1204511      | bacterioferritin comigratory protein [Haemophilus influenzae]   | 59    | 45      | 351         |
| 573       | 2      | 449        | 940       | gi 386681       | ORF YAL022 [Saccharomyces cerevisiae]   | 59    | 36      | 492         |
| 650       | 1      | 5          | 748       | gi 396400       | similar to eukaryotic Na/H <sup>+</sup> exchangers [Escherichia coli] sp P32703 YJCE_ECOLI HYPOTHETICAL 60.5 KD PROTEIN IN SOXR-ACS NTERGENIC REGION (0549) | 59    | 30      | 744         |
| 664       | 1      | 566        | 285       | gi 1263748      | lukF-PV like component [Staphylococcus aureus]  | 59    | 33      | 282         |
| 670       | 1      | 3          | 455       | gi 1123758      | unknown [Bacillus subtilis]   | 59    | 42      | 453         |
| 674       | 3      | 543        | 929       | gi 293033       | Integrase [Bacteriophage phi-LC3]   | 59    | 46      | 387         |
| 758       | 1      | 349        | 176       | gi 1500472      | M. jannaschii predicted coding region M21577 [Methanococcus jannaschii]   | 59    | 37      | 174         |
| 771       | 2      | 2270       | 1461      | gi 522150       | bromoperoxidase BPO-A1 [Streptomyces aureofaciens] sp P39112 BPAL-STRAU NON-HAEM BROMOPEROXIDASE BPO-A1 (EC 1.11.1.-) BROMIDE PEROXIDASE (BPO1).            | 59    | 44      | 810         |
| 825       | 1      | 2191       | 1097      | gi 397526       | clumping factor [Staphylococcus aureus]   | 59    | 47      | 1095        |
| 1032      | 2      | 1094       | 723       | gi 289262       | comE ORF3 [Bacillus subtilis]   | 59    | 36      | 372         |
| 1152      | 1      | 373        | 188       | gi 1276668      | ORF218 gene product [Porphyra purpurea]   | 59    | 37      | 186         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 1138      | 1      | 492        | 247       | gi1342419       | ATP-dependent nuclease [Bacillus subtilis]   | 59    | 26      | 246         |
| 1441      | 1      | 468        | 235       | gi13045942      | glycyl-tRNA synthetase [Mycoplasmag genitalium]  | 59    | 37      | 234         |
| 2103      | 1      | 1          | 186       | gi1459250       | triacylglycerol lipase [Galactomyces geotrichum]   | 59    | 33      | 186         |
| 2205      | 1      | 793        | 398       | gi1303794       | Yqem [Bacillus subtilis]   | 59    | 38      | 396         |
| 2578      | 1      | 484        | 284       | gi1258003       | insulin-like growth factor binding protein complex acid-labile ubunit [rat, liver, Peptide, 601 aa]  | 59    | 48      | 201         |
| 2967      | 2      | 145        | 348       | gi1212730       | Yqmk [Bacillus subtilis]   | 59    | 44      | 204         |
| 3012      | 1      | 3          | 248       | gi1773571       | neurofilament protein NF70 [Helix asperae]   | 59    | 31      | 246         |
| 3544      | 1      | 3          | 401       | gi11055218      | crotonase [Clostridium acetobutylicum]   | 59    | 42      | 399         |
| 3548      | 1      | 3          | 401       | gi11055218      | crotonase [Clostridium acetobutylicum]   | 59    | 42      | 399         |
| 3580      | 1      | 698        | 351       | gi11055218      | crotonase [Clostridium acetobutylicum]   | 59    | 42      | 348         |
| 3720      | 1      | 722        | 363       | gi11408494      | homologous to penicillin acylase [Bacillus subtilis]   | 59    | 36      | 360         |
| 4171      | 1      | 3          | 296       | gi11055218      | crotonase [Clostridium acetobutylicum]   | 59    | 42      | 294         |
| 4305      | 1      | 618        | 310       | gi11524193      | unknown [Mycobacterium tuberculosis]   | 59    | 39      | 309         |
| 18        | 1      | 1242       | 622       | gi1146913       | N-acetylglucosamine transport protein [Escherichia coli p1r1829895]MOEC2N phosphotransferase system enzyme I1 (EC 7.1.69), N-acetylglucosamine specific - Escherichia coli sp[p09123]PTAA-ECOLI PTS SYSTEM, N-ACETYLGLUCOSAMINE-SPECIFIC IABC COMPONENT [E11A] | 58    | 43      | 621         |
| 20        | 7      | 7020       | 5845      | gi150502        | collagen alpha chain precursor (AA-27 to 1127) [Mus musculus]  | 58    | 50      | 1176        |
| 21        | 5      | 3234       | 3626      | gi11054860      | phosphoribosyl anthranilate isomerase [Thermotoga maritima]  | 58    | 32      | 393         |
| 23        | 2      | 2841       | 1669      | gi11276880      | Epac [Streptococcus thermophilus]  | 58    | 29      | 1173        |
| 23        | 10     | 9301       | 8090      | gi11111133      | diaminopimelate decarboxylase (EC 4.1.1.20) - Pseudomonas aeruginosa   | 58    | 37      | 1212        |
| 38        | 29     | 22555      | 22884     | gi1973249       | vestitone reductase [Medicago sativa]  | 58    | 37      | 330         |
| 44        | 1      | 2          | 406       | gi1289272       | ferrichrome-binding protein [Bacillus subtilis]  | 58    | 33      | 405         |
| 45        | 1      | 1          | 552       | gi129464        | embryonic myosin heavy chain (1085 AA) [Homo sapiens] ir[S12460]S12460 myosin beta heavy chain - human   | 58    | 33      | 552         |
| 55        | 2      | 759        | 538       | gi1158852       | glucose regulated protein [Echinococcus multilocularis]  | 58    | 32      | 222         |
| 62        | 13     | 8493       | 8068      | gi1975353       | kinase-associated protein B [Bacillus subtilis]  | 58    | 35      | 426         |
| 63        | 3      | 1553       | 1717      | gi1166926       | [Arabidopsis thaliana unidentified mRNA sequence, complete cds.1, ene product [Arabidopsis thaliana]   | 58    | 35      | 165         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 67        | 13     | 12017      | 11229     | gi 1228083      | NADH dehydrogenase subunit 2 [Chortippus parallelus]                              | 58    | 41      | 789         |
| 96        | 8      | 8208       | 9167      | gi 709992       | hypothetical protein [Bacillus subtilis]  | 58    | 42      | 960         |
| 107       | 2      | 2065       | 1364      | gi 806327       | [Escherichia coli hcrA gene for A protein similar to yeast PEP16 and PEP22]       | 58    | 37      | 702         |
| 112       | 7      | 4519       | 5613      | gi 155588       | glucose-fructose oxidoreductase [Zymomonas mobilis] pfr[A42289]A42289             | 58    | 38      | 1095        |
| 114       | 6      | 7318       | 6503      | gi 1177843      | glucose-fructose oxidoreductase (EC 1.1.1.1) reductase - Zymomonas mobilis        | 58    | 38      | 816         |
| 140       | 2      | 2261       | 1395      | gi A45605 A456  | mature-parasite-infected erythrocyte surface antigen ME5A - Plasmodium falciparum | 58    | 31      | 867         |
| 151       | 2      | 717        | 950       | gi 1170261      | unknown [Mycobacterium tuberculosis]  | 58    | 31      | 234         |
| 154       | 6      | 8015       | 4827      | gi 1209277      | pCTH01 gene product [Chlamydia trachomatis]                                       | 58    | 41      | 1389        |
| 154       | 16     | 14281      | 13541     | gi 146613       | DNA ligase (EC 6.5.1.2) [Escherichia coli]  | 58    | 39      | 741         |
| 155       | 3      | 2269       | 1892      | gi 1303917      | Xq1B [Bacillus subtilis]  | 58    | 36      | 378         |
| 174       | 1      | 1056       | 529       | gi 904398       | hypothetical protein [Bacillus subtilis]  | 58    | 26      | 528         |
| 189       | 4      | 1533       | 1769      | gi 467383       | DNA binding protein (probable) [Bacillus subtilis]                                | 58    | 25      | 237         |
| 201       | 3      | 2669       | 3307      | gi 1511453      | endonuclease III [Methanococcus jannaschii]                                       | 58    | 34      | 639         |
| 204       | 1      | 2          | 238       | gi 1276729      | phycobilisome linker polypeptide [Porphyra purpurea]                              | 58    | 29      | 237         |
| 220       | 11     | 14575      | 13058     | gi 397526       | clumping factor [Staphylococcus aureus]   | 58    | 51      | 1518        |
| 231       | 3      | 1629       | 1474      | gi 1002520      | HutS [Bacillus subtilis]  | 58    | 45      | 156         |
| 233       | 6      | 4201       | 3497      | gi 1463023      | No definition line found [Caenorhabditis elegans]                                 | 58    | 39      | 705         |
| 243       | 10     | 9303       | 10082     | gi 537207       | ORF_277 [Escherichia coli]  | 58    | 32      | 780         |
| 257       | 1      | 331        | 1143      | gi 1340128      | ORF1 [Staphylococcus aureus]  | 58    | 44      | 813         |
| 302       | 2      | 460        | 801       | gi 40174        | ORF X [Bacillus subtilis]   | 58    | 34      | 342         |
| 307       | 11     | 6984       | 6127      | gi 1303842      | YqTU [Bacillus subtilis]  | 58    | 30      | 858         |
| 321       | 3      | 1914       | 2747      | gi 1239996      | hypothetical protein [Bacillus subtilis]  | 58    | 41      | 834         |
| 342       | 4      | 2724       | 3497      | gi 454838       | ORF 6: putative [Pseudomonas aeruginosa]  | 58    | 41      | 774         |
| 348       | 1      | 1          | 663       | gi 467478       | unknown [Bacillus subtilis]   | 58    | 36      | 663         |
| 401       | 2      | 384        | 605       | gi 143407       | para-aminobenzoic acid synthase, component I (pab) [Bacillus subtilis]            | 58    | 53      | 222         |

TABLE 2

St. aureus - putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 437       | 1      | 325        | 1554      | gi1303866       | YggS (Bacillus subtilis)  | 58    | 35      | 1230        |
| 445       | 1      | 105        | 1442      | gi1581583       | protein A (Staphylococcus aureus)   | 58    | 32      | 1338        |
| 453       | 3      | 789        | 965       | gi11009455      | unknown (Schizosaccharomyces pombe)   | 58    | 34      | 177         |
| 453       | 5      | 2748       | 2047      | gi1537214       | yjgC gene product (Escherichia coli)  | 58    | 40      | 702         |
| 479       | 2      | 731        | 1444      | gi11256621      | 26.7% of identity in 165 aa to a thermophilic bacterium hypothetical protein 6; putative (Bacillus subtilis)  | 58    | 36      | 714         |
| 490       | 1      | 909        | 547       | gi1580920       | rodd (gtaA) polypeptide (AA 1-673) (Bacillus subtilis) p1r1506048[S06048] probable rodd protein - Bacillus subtilis sp13184[TAGE_BACSU PROBABLE POLY(GLYCEROL-PHOSPHATE) LIPID-GLUCOSYLTRANSFERASE (EC 2.4.1.52) (TECHOIC ACID BIOSYNTHESIS PROTEIN E)] | 58    | 36      | 363         |
| 517       | 1      | 1          | 1164      | gi147264[V018]  | HYPOTHETICAL HELICASE HGO18   |       |         |             |
| 517       | 6      | 4182       | 4344      | gi1453422       | lor268 gene product (Mycoplasma hominis)  | 58    | 30      | 1164        |
| 546       | 3      | 2802       | 4019      | gi1886052       | restriction modification system S subunit (Spiroplasma citri) gi1886052 restriction modification system S subunit (Spiroplasma citri)   | 58    | 29      | 363         |
| 562       | 1      | 3          | 379       | gi143831        | nifS protein (AA 1-400) (Klebsiella pneumoniae)   | 58    | 37      | 1218        |
| 600       | 2      | 1347       | 1156      | gi11181839      | unknown (Pseudomonas aeruginosa)  | 58    | 34      | 177         |
| 604       | 2      | 1231       | 1001      | gi11001353      | hypothetical protein (Synecocystis sp.)   | 58    | 48      | 192         |
| 619       | 1      | 1          | 504       | gi1903748       | integral membrane protein (Homo sapiens)  | 58    | 41      | 231         |
| 625       | 1      | 2          | 364       | gi11208474      | hypothetical protein (Synecocystis sp.)   | 58    | 43      | 504         |
| 635       | 1      | 1492       | 755       | gi11510955      | transaldolase (Methanococcus jannaschii)  | 54    | 43      | 363         |
| 645       | 1      | 1          | 846       | gi1677882       | lileal sodium-dependent bile acid transporter (Rattus norvegicus) gi1677882 lileal sodium-dependent bile acid transporter (Rattus norvegicus)   | 58    | 41      | 738         |
| 645       | 3      | 906        | 1556      | gi11219899      | hypothetical protein (Bacillus subtilis)  | 58    | 33      | 846         |
| 665       | 1      | 771        | 532       | gi11204262      | hypothetical protein (DB10128-61) (Haemophilus influenzae)  | 58    | 41      | 651         |
| 674       | 1      | 635        | 327       | gi1498817       | ORF8; homologous to small subunit of phage terminases (Bacillus ubtilis)  | 58    | 39      | 240         |
| 675       | 2      | 1312       | 806       | gi142181        | osmc gene product (Escherichia coli)  | 58    | 39      | 309         |
| 745       | 1      | 618        | 310       | gi11205432      | coenzyme PQQ synthesis protein III (pqgIII) (Haemophilus influenzae)  | 58    | 28      | 507         |
| 799       | 2      | 242        | 1174      | gi11204669      | collagenase (Haemophilus influenzae)  | 58    | 32      | 309         |
| 800       | 2      | 1096       | 614       | gi1171963       | tRNA isopentenyl transferase (Saccharomyces cerevisiae) sp1607884[MOD5-YEAST TRNA ISOPENTENYLTRANSFERASE (EC 2.5.1.8) ISOPENTENYL-DIPHOSPHATE: TRNA ISOPENTENYLTRANSFERASE (IPP TRANSFERASE) (IPPT)]  | 58    | 36      | 933         |
|           |        |            |           |                 |   | 58    | 37      | 483         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 854       | 1      | 1108       | 805       | gi 466778       | lysine specific permease [Escherichia coli]  | 58    | 44      | 504         |
| 885       | 1      | 481        | 242       | gi 861199       | protoporphyrin IX Mg-chelatase subunit precursor [Hordium vulgare]   | 58    | 33      | 240         |
| 891       | 1      | 3          | 527       | gi 1293660      | AbsA2 [Streptomyces coelicolor]  | 58    | 31      | 525         |
| 942       | 1      | 931        | 467       | gi 405567       | trmH [Plasmodium falciparum]   | 58    | 30      | 465         |
| 1002      | 1      | 952        | 521       | gi 577649       | proLUMK [Staphylococcus aureus]  | 58    | 34      | 432         |
| 1438      | 1      | 1          | 261       | gi 583558       | isoleucyl tRNA synthetase [Staphylococcus aureus] sp P41368 SY/P_STANU ISOLEUCYL-TRNA SYNTHETASE, MUPROCIIN RESISTANT EC 6.1.1.51 (ISOLEUCINE-TRNA LIGASE) (ILERS) (MUPROCIIN RESISTANCE PROTEIN)    | 58    | 30      | 261         |
| 1442      | 1      | 2          | 463       | gi 971394       | similar to Acc No. D26185 [Escherichia coli]   | 58    | 34      | 462         |
| 1873      | 1      | 480        | 241       | gi 1131951      | small subunit of NADH-dependent glutamate synthase [Plectononema boryanum]   | 58    | 38      | 240         |
| 1876      | 1      | 3          | 158       | gi 529216       | No definition line found [Caenorhabditis elegans] sp P46503 YLM7_CAEEL HYPOTHETICAL 7.3 NO PROTEIN P2312.7 IN PRODOMONE III  | 58    | 33      | 156         |
| 1989      | 1      | 108        | 401       | gi 1405458      | YnfK [Bacillus subtilis]   | 58    | 29      | 294         |
| 2109      | 1      | 3          | 401       | gi 1001801      | hypothetical protein [Synecocystis sp.]  | 58    | 31      | 399         |
| 2473      | 1      | 288        | 145       | gi 510140       | oligoendopeptidase F [Lactococcus lactis]  | 58    | 38      | 144         |
| 2523      | 1      | 452        | 228       | gi 644873       | catabolic dehydroquinase dehydratase [Acinetobacter calcoaceticus]   | 58    | 37      | 225         |
| 3041      | 1      | 2          | 211       | gi 1205367      | oligopeptide transport ATP-binding protein [Haemophilus influenzae]  | 58    | 39      | 210         |
| 3094      | 1      | 3          | 263       | gi 1185288      | isochorismate synthase [Bacillus subtilis]   | 58    | 38      | 261         |
| 3706      | 1      | 3          | 383       | gi 456614       | mevalonate kinase [Arabidopsis thaliana]   | 58    | 48      | 381         |
| 3854      | 1      | 1          | 402       | gi 808869       | human gcp372 [Homo sapiens]  | 58    | 32      | 402         |
| 4082      | 1      | 51         | 224       | gi 508551       | ribulose-1,5 biphosphate carboxylase large subunit -methyltransferase [Pisum sativum]  | 58    | 37      | 174         |
| 4278      | 1      | 3          | 206       | gi 180189       | cerebellar-degeneration-related antigen (CDR34) [Homo sapiens] gi 182737 cerebellar degeneration-associated protein [Homo sapiens] pif A29770 A29770 cerebellar degeneration-related protein - human | 58    | 37      | 204         |
| 19        | 7      | 7818       | 7363      | gi 1001516      | hypothetical protein [Synecocystis sp.]  | 57    | 31      | 456         |
| 23        | 11     | 9663       | 8872      | gi 1606066      | ORF_256 [Escherichia coli]   | 57    | 29      | 792         |
| 31        | 1      | 4801       | 2402      | gi 153116       | ORF3 [Streptomyces coelicolor]   | 57    | 32      | 2400        |
| 38        | 14     | 11611      | 10796     | gi 144859       | ORF B [Clostridium perfringens]  | 57    | 31      | 816         |
| 46        | 14     | 12063      | 12046     | gi 1001319      | hypothetical protein [Synecocystis sp.]  | 57    | 25      | 984         |

TABLE 2

5 aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 51        | 3      | 1411       | 1187      | pir B33856 B338 | hypothetical 80K protein - <i>Bacillus sphaericus</i>   | 57    | 38      | 225         |
| 54        | 1      | 1          | 453       | gi 564950       | staphylococcal accessory regulator A [ <i>Staphylococcus aureus</i> ]   | 57    | 31      | 453         |
| 75        | 1      | 3          | 239       | gi 1000470      | C27B7.7 [ <i>Caenorhabditis elegans</i> ]   | 57    | 42      | 237         |
| 92        | 5      | 3855       | 3061      | gi 143607       | sporulation protein [ <i>Bacillus subtilis</i> ]  | 57    | 35      | 795         |
| 96        | 3      | 4006       | 4773      | gi 144297       | acetyl esterase (XyNC) [ <i>Caldocellum saccharolyticum</i> ] pir B37202 B37202<br>acetyl esterase (EC 3.1.1.6) (XyNC) - <i>Caldocellum saccharolyticum</i> | 57    | 34      | 768         |
| 107       | 3      | 1480       | 2076      | gi 460955       | TagE [ <i>Vibrio cholerae</i> ]   | 57    | 42      | 597         |
| 109       | 8      | 5340       | 5933      | gi 1438846      | Unknown [ <i>Bacillus subtilis</i> ]  | 57    | 41      | 594         |
| 112       | 9      | 6679       | 7701      | gi 1486250      | Unknown [ <i>Bacillus subtilis</i> ]  | 57    | 33      | 1023        |
| 114       | 4      | 6384       | 4108      | gi 871456       | putative alpha subunit of formate dehydrogenase [ <i>Methanobacterium thermoautotrophicum</i> ]   | 57    | 37      | 2277        |
| 126       | 2      | 430        | 1053      | gi 288301       | ORF2 gene product [ <i>Bacillus megaterium</i> ]  | 57    | 37      | 624         |
| 131       | 5      | 6537       | 6277      | gi 1511160      | M. jannaschii predicted coding region M01163 [ <i>Methanococcus jannaschii</i> ]  | 57    | 38      | 261         |
| 133       | 3      | 2668       | 2201      | gi 1303912      | Yqhw [ <i>Bacillus subtilis</i> ]   | 57    | 40      | 468         |
| 133       | 4      | 3383       | 2784      | gi 1221884      | [urea?] amidolyase [ <i>Haemophilus influenzae</i> ]  | 57    | 37      | 600         |
| 147       | 4      | 2164       | 1694      | gi 467469       | Unknown [ <i>Bacillus subtilis</i> ]  | 57    | 33      | 471         |
| 160       | 2      | 1293       | 1060      | gi 558604       | chitin synthase 2 [ <i>Neurospora crassa</i> ]  | 57    | 28      | 234         |
| 163       | 8      | 5687       | 4764      | gi 145580       | icarD gene product [ <i>Escherichia coli</i> ]  | 57    | 38      | 924         |
| 168       | 6      | 4336       | 5325      | gi 39782        | 33kDa lipoprotein [ <i>Bacillus subtilis</i> ]  | 57    | 32      | 990         |
| 170       | 5      | 3297       | 3455      | gi 603404       | Ver164p [ <i>Saccharomyces cerevisiae</i> ]   | 57    | 37      | 159         |
| 221       | 6      | 8026       | 6809      | gi 1136221      | carboxypeptidase [ <i>Sulfolobus solfataricus</i> ]   | 57    | 32      | 1218        |
| 224       | 3      | 1348       | 1791      | gi 268969       | fibronectin binding protein [ <i>Streptococcus dysgalactiae</i> ] pir S33850 S33850<br>fibronectin-binding protein - <i>Streptococcus dysgalactiae</i>      | 57    | 32      | 444         |
| 263       | 4      | 4411       | 3886      | gi 1185002      | dihydrodipicolinate reductase [ <i>Pseudomonas syringae</i> pv. <i>tabaci</i> ]   | 57    | 42      | 726         |
| 276       | 1      | 494        | 255       | gi 396380       | No definition line found [ <i>Escherichia coli</i> ]  | 57    | 40      | 240         |
| 283       | 2      | 335        | 1324      | gi 773349       | B1A protein [ <i>Bacillus subtilis</i> ]  | 57    | 32      | 990         |
| 297       | 1      | 469        | 236       | gi 1334820      | reading frame V [ <i>Cauliflower mosaic virus</i> ]   | 57    | 46      | 234         |
| 342       | 3      | 1993       | 2805      | gi 1204431      | hypothetical protein (SP:P33644) [ <i>Haemophilus influenzae</i> ]  | 57    | 35      | 813         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 375       | 6      | 3340       | 3741      | gi 385177       | cell division protein [Bacillus subtilis]   | 57    | 26      | 402         |
| 433       | 6      | 3286       | 4011      | gi 1524117      | alpha-acetolactate decarboxylase [Lactococcus lactis]   | 57    | 40      | 726         |
| 470       | 3      | 903        | 1145      | gi 804819       | protein serine/threonine kinase [Toxoplasma gondii]   | 57    | 30      | 243         |
| 487       | 5      | 1391       | 1723      | gi 507323       | ORF1 [Bacillus stearothermophilus]  | 57    | 28      | 333         |
| 498       | 1      | 274        | 852       | gi 1134549      | NADH-ubiquinone oxidoreductase subunit 4L [Podospora anserina]  | 57    | 34      | 579         |
| 503       | 1      | 343        | 173       | gi 1502283      | organic cation transporter OCT2 [Rattus norvegicus]   | 57    | 30      | 171         |
| 505       | 2      | 1619       | 1284      | gi 466886       | glc96_C2_194 [Mycobacterium leprae]   | 57    | 40      | 336         |
| 519       | 2      | 1182       | 2549      | gi 1303707      | Yrzh [Bacillus subtilis]  | 57    | 34      | 1368        |
| 522       | 2      | 3234       | 1945      | gi 1064809      | homologous to sp:MPRA_ECOLI [Bacillus subtilis]   | 57    | 36      | 1290        |
| 538       | 2      | 909        | 1415      | gi 153179       | phosphorothiohyacin n-acetyltransferase [Streptomyces coelicolor] p1rj00246/p10246 phosphothiohyacin N-acetyltransferase (EC 2.3.1.-) Streptomyces coelicolor | 57    | 40      | 507         |
| 547       | 1      | 968        | 486       | gi 467340       | unknown [Bacillus subtilis]   | 57    | 50      | 483         |
| 599       | 1      | 1062       | 532       | sp P20692 TYRA  | PREPHENATE DEHYDROGENASE (EC 1.3.1.12) (PDH)  | 57    | 41      | 531         |
| 630       | 2      | 757        | 572       | gi 1107894      | unknown [Schizosaccharomyces pombe]   | 57    | 38      | 186         |
| 622       | 2      | 1600       | 1130      | gi 173028       | thioredoxin 11 [Saccharomyces cerevisiae]   | 57    | 39      | 471         |
| 625       | 2      | 362        | 1114      | gi 1262366      | hypothetical protein [Mycobacterium leprae]   | 57    | 34      | 753         |
| 680       | 1      | 1          | 204       | gi 143544       | RNA polymerase sigma-30 factor [Bacillus subtilis] p1rj28025/p128025 transcription initiation factor sigma H - cellus subtilis                                | 57    | 30      | 204         |
| 690       | 1      | 3          | 629       | gi 466520       | pocr [Salmonella typhimurium]   | 57    | 29      | 627         |
| 696       | 1      | 2          | 433       | gi 413972       | ipa-48r gene product [Bacillus subtilis]  | 57    | 33      | 432         |
| 704       | 1      | 36         | 638       | gi 1699331      | M. jannaschii predicted coding region M2100 [Methanococcus jannaschii]  | 57    | 36      | 603         |
| 732       | 1      | 2316       | 1621      | gi 1418999      | orf4 [Lactobacillus sake]   | 57    | 37      | 696         |
| 746       | 1      | 451        | 227       | gi 392973       | Ran3 [Aplysia californica]  | 57    | 42      | 225         |
| 757       | 1      | 20         | 466       | gi 43979        | L. curvatus small cryptic plasmid gene for rep protein [Lactobacillus rvaus]  | 57    | 45      | 447         |
| 862       | 1      | 2          | 295       | gi 1303827      | Yqf1 [Bacillus subtilis]  | 57    | 21      | 294         |
| 1049      | 1      | 907        | 435       | gi 1510108      | ORF-1 [Agrobacterium tumefaciens]   | 57    | 35      | 453         |
| 1117      | 1      | 1387       | 695       | gi 896286       | MW2 terminus uncertain [Leishmania tarentolae]  | 57    | 28      | 693         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 1136      | 1      | 2          | 322       | gi1303853       | YqgF (Bacillus subtilis)   | 57    | 38      | 321         |
| 1144      | 2      | 1033       | 611       | gi1310083       | voltege-activated calcium channel alpha-1 subunit [Rattus oryzeolus]                                     | 57    | 46      | 423         |
| 1172      | 1      | 1472       | 738       | gi1511146       | M. jannaschii predicted coding region MJ1143 [Methanococcus jannaschii]                                  | 57    | 28      | 735         |
| 1500      | 2      | 746        | 558       | gi142780        | putative membrane protein; putative [Bacillus subtilis]  | 57    | 35      | 189         |
| 1676      | 1      | 659        | 399       | gi1313777       | ureacil permease [Escherichia coli]  | 57    | 31      | 261         |
| 2481      | 1      | 2          | 400       | gi1237015       | ORF4 [Bacillus subtilis]   | 57    | 23      | 399         |
| 3099      | 1      | 3          | 230       | gi1204540       | isochorismate synthase [Haemophilus influenzae]  | 57    | 19      | 224         |
| 3122      | 1      | 360        | 181       | gi1882472       | ORF_0464 [Escherichia coli]  | 57    | 40      | 180         |
| 3560      | 1      | 2          | 361       | gi153490        | tetracycline C resistance and export protein [Streptomyces laevis]                                       | 57    | 37      | 360         |
| 3850      | 1      | 856        | 434       | gi1155588       | glucose-fructose oxidoreductase [Zymomonas mobilis] pifA42289/A42289                                     | 57    | 40      | 423         |
| 3931      | 1      | 704        | 354       | gi1413953       | glucose-fructose oxidoreductase [EC 1.1.1.88] recursor - Zymomonas mobilis                               | 57    | 36      | 351         |
| 3993      | 1      | 1          | 384       | gi1151259       | lpa-29d gene product [Bacillus subtilis]   | 57    | 39      | 384         |
| 4065      | 1      | 793        | 398       | gi1300037       | UMP-CoA reductase [EC 1.1.1.88] [Pseudomonas nevalonii] pifA44756/A44756                                 | 57    | 31      | 396         |
| 4100      | 1      | 596        | 300       | gi11086633      | hydroxymethylglutaryl-CoA reductase [EC 1.1.1.88] Pseudomonas sp.  | 57    | 47      | 297         |
| 4163      | 1      | 571        | 287       | gi131512        | nitrate reductase [EC 1.7.99.4] alpha chain - Escherichia coli   | 57    | 50      | 284         |
| 4267      | 2      | 631        | 335       | gi11000365      | T06C10.5 gene product [Caenorhabditis elegans]   | 57    | 38      | 297         |
| 4358      | 1      | 3          | 302       | gi1298032       | [patatin [Solanum tuberosum]   | 57    | 32      | 300         |
| 4389      | 2      | 108        | 290       | gi1405894       | SpoIIAG [Bacillus subtilis]  | 57    | 37      | 183         |
| 4399      | 1      | 2          | 232       | gi11483603      | EF [Streptococcus suis]  | 57    | 35      | 231         |
| 4481      | 1      | 572        | 288       | gi1405879       | 1-phosphofructokinase [Escherichia coli]   | 57    | 44      | 285         |
| 4486      | 1      | 512        | 258       | gi1515938       | [pristinamycin I synthase I [Streptomyces pristinaespiralis]   | 57    | 42      | 255         |
| 4510      | 1      | 481        | 242       | gi1205301       | yeiH [Escherichia coli]  | 57    | 38      | 240         |
| 4617      | 1      | 468        | 256       | gi1511222       | glutamate synthase (ferredoxin) [Synecocystis sp.] pifA46957/A46957                                      | 57    | 35      | 213         |
| 4         | 31     | 12201      | 11524     | gi1149204       | glutamate synthase (ferredoxin) [EC 1.4.7.1] - Synecocystis sp.  | 56    | 31      | 678         |
|           |        |            |           |                 | leukotoxin secretion ATP-binding protein [Haemophilus influenzae]  |       |         |             |
|           |        |            |           |                 | restriction modification enzyme, subunit M1 [Methanococcus jannaschii]                                   |       |         |             |
|           |        |            |           |                 | histidine utilization repressor C [Klebsiella aerogenes] pifA36730/A36730                                |       |         |             |
|           |        |            |           |                 | huto protein - Klebsiella pneumoniae (fragment) sp P19452 HUTO_KLEAE                                     |       |         |             |
|           |        |            |           |                 | FORMINOGUTAMINASE [EC 3.5.3.8] FORMINOGUTAMINASE (HYDROLASE) (HISTIDINE UTILIZATION PROTEIN G) FRAGMENT) |       |         |             |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Cont'g | ORF  | Start | Stop  | match      | match   | match | % sim | % ident | length |
|--------|------|-------|-------|------------|---|-------|-------|---------|--------|
| ID     | (nt) | (nt)  | (nt)  | accession  | gene name   |       |       |         | (nt)   |
| 22     | 8    | 4248  | 5177  | gi 1322222 | RACH1 [Homo sapiens]  |       | 56    | 33      | 930    |
| 36     | 28   | 21179 | 22264 | gi 1480705 | lipote-protein ligase [Mycoplasma capricolum]   |       | 56    | 34      | 1086   |
| 44     | 3    | 1861  | 2421  | gi 450320  | Y gene product (unidentified)   |       | 56    | 31      | 561    |
| 44     | 15   | 10103 | 10606 | gi 1203099 | hypothetical protein (G8:U19201.1) [Haemophilus influenzae]   |       | 56    | 39      | 504    |
| 50     | 6    | 4820  | 5161  | gi 209331  | fiber protein [Human adenovirus type 5]   |       | 56    | 48      | 342    |
| 53     | 4    | 2076  | 2372  | gi 623476  | transcriptional activator [Providencia stuartii] sp P43463 AARP_PROST<br>TRANSCRIPTIONAL ACTIVATOR AARP |       | 56    | 30      | 897    |
| 67     | 6    | 5656  | 6594  | gi 466613  | nika [Escherichia coli]   |       | 56    | 32      | 939    |
| 89     | 3    | 2164  | 1810  | gi 482922  | protein with homology to pail repressor of B. subtilis [Lactobacillus elbructii]                        |       | 56    | 39      | 555    |
| 96     | 1    | 203   | 913   | gi 145594  | cAMP receptor protein (crp) [Escherichia coli]  |       | 56    | 35      | 711    |
| 109    | 21   | 18250 | 17846 | gi 1204367 | hypothetical protein (G8:U14003_278) [Haemophilus influenzae]   |       | 56    | 27      | 405    |
| 112    | 8    | 5611  | 6678  | gi 155588  | glucose-fructose oxidoreductase (Glc 1.1.-.) reductase - Zymomonas mobilis                              |       | 56    | 40      | 1068   |
| 131    | 3    | 6404  | 5100  | gi 619724  | MgtE [Bacillus firmus]  |       | 56    | 30      | 1305   |
| 136    | 2    | 65    | 232   | gi 413948  | ipa-24d gene product [Bacillus subtilis]  |       | 56    | 31      | 168    |
| 138    | 4    | 823   | 1521  | gi 580868  | ipa-22r gene product [Bacillus subtilis]  |       | 56    | 31      | 699    |
| 146    | 2    | 740   | 447   | gi 1046009 | M. genitalium predicted coding region MG309 [Mycoplasma genitalium]                                     |       | 56    | 37      | 294    |
| 149    | 2    | 1639  | 1067  | gi 945380  | terminase small subunit [Bacteriophage LU-H]  |       | 56    | 35      | 573    |
| 163    | 1    | 2     | 223   | gi 143947  | glucamine synthetase [Bacteroides fragilis]   |       | 56    | 30      | 223    |
| 166    | 5    | 6745  | 6449  | gi 405792  | ORF154 [Pseudomonas putida]   |       | 56    | 26      | 297    |
| 187    | 1    | 31    | 393   | gi 311237  | H(+)-transporting ATP synthase [Zea mays]   |       | 56    | 30      | 363    |
| 190    | 1    | 2     | 373   | gi 1109686 | ProX [Bacillus subtilis]  |       | 56    | 35      | 372    |
| 191    | 8    | 11516 | 9943  | gi 581070  | acyl coenzyme A synthetase [Escherichia coli]   |       | 56    | 35      | 1596   |
| 195    | 3    | 1291  | 647   | gi 1510242 | collagenase [Methanococcus jannaschii]  |       | 56    | 34      | 645    |
| 210    | 3    | 2323  | 2072  | gi 40163   | heat shock protein [Clostridium acetobutylicum]   |       | 56    | 39      | 252    |
| 218    | 5    | 3383  | 3775  | gi 147533  | sarA [Staphylococcus aureus]  |       | 56    | 31      | 383    |
| 270    | 2    | 813   | 1712  | gi 765073  | autolysin [Staphylococcus aureus]   |       | 56    | 41      | 900    |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 290       | 1      | 3221       | 1632      | gi1547513       | orf3 [Haemophilus influenzae]   | 56    | 34      | 1590        |
| 297       | 5      | 1140       | 1373      | gi1511556       | M. jannaschii predicted coding region MJ1561 [Methanococcus jannaschii]                                       | 56    | 40      | 234         |
| 321       | 2      | 2947       | 1799      | gi1001801       | hypothetical protein [Synecocystis sp.]   | 56    | 31      | 1149        |
| 359       | 2      | 1279       | 641       | gi146336        | ncif gene product [Rhizobium meliloti]  | 56    | 26      | 639         |
| 371       | 2      | 360        | 1823      | gi145304        | U-ribulokinase [Escherichia coli]   | 56    | 39      | 1464        |
| 391       | 4      | 1762       | 2409      | gi1001634       | hypothetical protein [Synecocystis sp.]   | 56    | 34      | 648         |
| 402       | 1      | 380        | 192       | gi1428904       | [S-HT4] receptor [Homo sapiens]   | 56    | 48      | 189         |
| 416       | 4      | 2480       | 2109      | gi1408486       | HS74A gene product [Bacillus subtilis]  | 56    | 31      | 372         |
| 424       | 3      | 1756       | 2334      | gi142471        | acetolactate decarboxylase [Bacillus subtilis]  | 56    | 32      | 579         |
| 457       | 1      | 1907       | 1017      | gi1205194       | [formamidopyrimidine-DNA glycosylase [Haemophilus influenzae]   | 56    | 36      | 891         |
| 458       | 2      | 2423       | 1812      | gi15466         | terminase [Bacteriophage SP1]   | 56    | 37      | 612         |
| 504       | 2      | 2152       | 1283      | gi1142681       | [lpp38 [Pasteurella haemolytica]  | 56    | 38      | 870         |
| 511       | 1      | 1          | 1284      | gi1217049       | [brnQ protein [Salmonella typhimurium]  | 56    | 37      | 1284        |
| 604       | 3      | 1009       | 1701      | gi1467109       | rlm: 30S ribosomal protein S18 alanine acetyltransferase: 229_C1_170 [Mycobacterium leprae]                   | 56    | 43      | 603         |
| 660       | 5      | 3547       | 3774      | gi1229106       | [ZK910.1 [Caenorhabditis elegans]   | 56    | 30      | 228         |
| 707       | 1      | 35         | 400       | gi155929        | [NADPH-sulfite reductase flavoprotein component [Salmonella typhimurium]                                      | 56    | 38      | 366         |
| 709       | 2      | 1385       | 1095      | gi1510801       | hydrogenase accessory protein [Methanococcus jannaschii]  | 56    | 38      | 291         |
| 718       | 1      | 1          | 495       | gi1413948       | [ipa-24d gene product [Bacillus subtilis]   | 56    | 35      | 495         |
| 743       | 1      | 87         | 677       | gi1928636       | [repressor protein [Lactococcus lactis phage BK5-7]   | 56    | 35      | 591         |
| 790       | 1      | 726        | 399       | gi1511513       | ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii]                                      | 56    | 33      | 378         |
| 795       | 1      | 3          | 407       | gi1205382       | [cell division protein [Haemophilus influenzae]   | 56    | 34      | 405         |
| 813       | 1      | 19         | 930       | gi1222161       | [permease [Haemophilus influenzae]  | 56    | 28      | 912         |
| 855       | 1      | 3          | 515       | gi1256621       | [26.7% of identity in 165 aa to a Thermophilic bacterium hypothetical protein 6; putative [Bacillus subtilis] | 56    | 33      | 513         |
| 968       | 1      | 2          | 466       | gi1547513       | orf3 [Haemophilus influenzae]   | 56    | 37      | 465         |
| 973       | 2      | 1049       | 732       | gi1864022       | [Hear [Pseudomonas aeruginosa]  | 56    | 71      | 318         |
| 1203      | 1      | 5          | 223       | gi184251        | [HMG-1 [Homo sapiens]   | 56    | 34      | 219         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 1976      | 1      | 452        | 237       | gi 19806        | lysine-rich aspartic acid-rich protein [Plasmodium chabaudi]<br>r[S22183]S22183 lysine/aspartic acid-rich protein - Plasmodium baudii             | 56    | 33      | 216         |
| 2161      | 1      | 2          | 400       | gi 1237015      | ORF4 [Bacillus subtilis]  | 56    | 27      | 399         |
| 2258      | 1      | 362        | 183       | gi 466685       | No definition line found [Escherichia coli]   | 56    | 26      | 180         |
| 2279      | 1      | 421        | 212       | gi 1204354      | spore germination and vegetative growth protein [Haemophilus influenzae]  | 56    | 40      | 210         |
| 2294      | 2      | 526        | 326       | gi 1036646      | phosphoribosylformino-prate ketoisomerase [Rhodobacter phaeoideus]  | 56    | 29      | 201         |
| 3026      | 1      | 179        | 328       | gi 143306       | penicillin V amidase [Bacillus sphaericus]  | 56    | 30      | 150         |
| 3189      | 1      | 289        | 146       | gi 1186604      | Similar to aldehyde dehydrogenase [Caenorhabditis elegans]  | 56    | 37      | 144         |
| 3770      | 1      | 63         | 401       | gi 1129145      | acetyl-CoA C-acyltransferase [Mangifera indica]   | 56    | 43      | 339         |
| 4054      | 2      | 720        | 361       | gi 1205355      | Na+/H+ antiporter [Haemophilus influenzae]  | 56    | 31      | 360         |
| 4145      | 1      | 1          | 324       | gi 126095       | long-chain acyl-CoA dehydrogenase [Mus musculus]  | 56    | 36      | 324         |
| 4200      | 1      | 505        | 254       | gi 155588       | glucose-fructose oxidoreductase [Zymomonas mobilis] p1r[A2289]A2289<br>glucose-fructose oxidoreductase (EC 1.1.1.-) precursor - Zymomonas mobilis | 56    | 40      | 252         |
| 4273      | 1      | 675        | 355       | gi 308861       | OTC start codon [Lactococcus lactis]  | 56    | 33      | 321         |
| 1         | 3      | 4095       | 3436      | gi 5341         | Putative orf YCLM8c, len:192 [Saccharomyces cerevisiae] r[S53591]S53591<br>hypothetical protein - yeast [Saccharomyces cerevisiae]                | 55    | 25      | 660         |
| 11        | 12     | 9377       | 8505      | gi 216773       | haloacetate dehalogenase H-1 [Moraxella sp.]  | 55    | 32      | 873         |
| 12        | 4      | 5133       | 4534      | gi 467337       | unknown [Bacillus subtilis]   | 55    | 26      | 600         |
| 19        | 5      | 5404       | 5844      | gi 1001719      | hypothetical protein [Synecocystis sp.]   | 55    | 25      | 441         |
| 23        | 13     | 14087      | 12339     | gi 474190       | lucA gene product [Escherichia coli]  | 55    | 30      | 1749        |
| 32        | 7      | 5368       | 6888      | gi 1340056      | unknown [Mycobacterium tuberculosis]  | 55    | 37      | 1521        |
| 34        | 3      | 2569       | 1808      | gi 1303968      | YqjQ [Bacillus subtilis]  | 55    | 39      | 762         |
| 34        | 5      | 3960       | 3412      | gi 1303962      | YqjK [Bacillus subtilis]  | 55    | 33      | 549         |
| 36        | 1      | 1291       | 647       | gi 1606045      | ORF_0118 [Escherichia coli]   | 55    | 27      | 645         |
| 36        | 6      | 6220       | 5243      | gi 1001341      | hypothetical protein [Synecocystis sp.]   | 55    | 31      | 978         |
| 47        | 3      | 3054       | 3821      | gi 1001819      | hypothetical protein [Synecocystis sp.]   | 55    | 21      | 768         |
| 49        | 1      | 2065       | 1127      | gi 403373       | glycerophosphoryl diester phosphodiesterase [Bacillus subtilis]<br>p1r[S3725]S3725 glycerophosphoryl diester phosphodiesterase - acillus subtilis | 55    | 36      | 939         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 67        | 11     | 8966       | 9565      | [gi1153053]     | norA199 protein [Staphylococcus aureus]   | 55    | 23      | 600         |
| 75        | 3      | 881        | 1273      | [gi121698]      | [L-histidinol: NAD <sup>+</sup> oxidoreductase (EC 1.1.1.23) (aa 1-434) Escherichia coli] | 55    | 33      | 393         |
| 82        | 9      | 11587      | 14194     | [gi1136221]     | carboxypeptidase [Sulfolobus solfataricus]  | 55    | 35      | 3194        |
| 87        | 4      | 3517       | 4917      | [gi11064812]    | function unknown [Bacillus subtilis]  | 55    | 28      | 1401        |
| 88        | 2      | 1172       | 1636      | [gi1882463]     | protein-Nip1-phosphohistidine-sugar phosphotransferase [Escherichia coli]                 | 55    | 35      | 465         |
| 92        | 1      | 127        | 516       | [gi11377832]    | unknown [Bacillus subtilis]   | 55    | 36      | 390         |
| 100       | 2      | 836        | 2035      | [gi11370274]    | leaxanthin epoxidase [Nicotiana glauca (folia)]   | 55    | 36      | 1200        |
| 100       | 5      | 5137       | 4658      | [gi1396660]     | unknown open reading frame [Buchnera aphidicola]  | 55    | 29      | 480         |
| 104       | 3      | 4266       | 2986      | [gi1149866]     | [M. jannaschii predicted coding region MJ1024 (Methanococcus jannaschii)]                 | 55    | 31      | 1281        |
| 114       | 3      | 2616       | 1834      | [gi11511167]    | [formate dehydrogenase, alpha subunit (Methanococcus jannaschii)]                         | 55    | 29      | 783         |
| 144       | 3      | 1805       | 1476      | [gi1100787]     | unknown [Saccharomyces cerevisiae]  | 55    | 35      | 330         |
| 165       | 5      | 6212       | 5508      | [gi1045884]     | [M. genitalium predicted coding region MG199 (Mycoplasma genitalium)]                     | 55    | 27      | 705         |
| 189       | 5      | 2205       | 2576      | [gi1142569]     | [ATP synthase a subunit (Bacillus firmus)]  | 55    | 35      | 372         |
| 191       | 6      | 9136       | 6857      | [gi1559411]     | [B0272.3 (Caenorhabditis elegans)]  | 55    | 39      | 2280        |
| 194       | 2      | 364        | 636       | [gi1145768]     | [K7 kinesin-like protein (Dictyostelium discoideum)]                                      | 55    | 34      | 273         |
| 209       | 4      | 1335       | 1676      | [gi1473357]     | [chi4 gene product (Schizosaccharomyces pombe)]   | 55    | 35      | 342         |
| 211       | 2      | 1693       | 1145      | [gi1410130]     | [ORF6 (Bacillus subtilis)]  | 55    | 37      | 549         |
| 213       | 2      | 644        | 1372      | [gi1633692]     | [TraA (Yersinia enterocolitica)]  | 55    | 28      | 729         |
| 214       | 7      | 4144       | 5481      | [gi11001793]    | hypothetical protein [Synecocystis sp.]   | 55    | 30      | 1338        |
| 221       | 7      | 11473      | 9197      | [gi1466520]     | [pocR (Salmonella typhimurium)]   | 55    | 32      | 2277        |
| 233       | 8      | 5908       | 4817      | [gi1237063]     | unknown [Mycobacterium tuberculosis]  | 55    | 38      | 1092        |
| 236       | 4      | 1375       | 2340      | [gi1146199]     | [putative (Bacillus subtilis)]  | 55    | 32      | 966         |
| 243       | 2      | 380        | 1885      | [gi1459907]     | [mercuric reductase (plasmid p1258)]  | 55    | 29      | 1506        |
| 258       | 1      | 786        | 394       | [gi1453006]     | [orf6 (Rhodococcus fascians)]   | 55    | 36      | 393         |
| 281       | 1      | 126        | 938       | [gi11408493]    | [homologous to SwissProt:YIDA_ECOLI hypothetical protein (Bacillus subtilis)]             | 55    | 35      | 813         |
| 316       | 3      | 1323       | 2102      | [gi11486447]    | [LysA homologue (Rhizobium sp.)]  | 55    | 30      | 780         |
| 326       | 5      | 2968       | 2744      | [gi11296824]    | [proline iminopeptidase (Lactobacillus halveticus)]                                       | 55    | 36      | 235         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Putative gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 351       | 2      | 2322       | 1429      | [gi11204820]    | hydrogen peroxide-inducible activator [Haemophilus influenzae]  | 55    | 28      | 894         |
| 353       | 4      | 2197       | 2412      | [gi11272475]    | chitin synthase [Emericella nidulans]   | 55    | 50      | 216         |
| 380       | 1      | 14         | 379       | [gi1142554]     | ATP synthase 1 subunit [Bacillus megaterium]  | 55    | 37      | 366         |
| 383       | 1      | 462        | 232       | [gi1149272]     | ferrichrome-binding protein [Bacillus subtilis]   | 55    | 36      | 231         |
| 386       | 1      | 3          | 938       | [gi11510251]    | DNA helicase, putative [Methanococcus jannaschii]   | 55    | 30      | 936         |
| 410       | 2      | 1208       | 1891      | [gi11205144]    | multidrug resistance protein [Haemophilus influenzae]   | 55    | 27      | 684         |
| 483       | 2      | 411        | 833       | [gi1113934]     | lpa-10r gene product [Bacillus subtilis]  | 55    | 26      | 423         |
| 529       | 3      | 1777       | 1433      | [gi1606150]     | ORF_3209 [Escherichia coli]   | 55    | 33      | 345         |
| 555       | 1      | 1088       | 585       | [gi1143407]     | para-aminobenzoic acid synthase, component 1 (pab) [Bacillus subtilis]  | 55    | 28      | 504         |
| 565       | 1      | 402        | 202       | [gi11223961]    | CDP-tylase epimerase [Yersinia pseudotuberculosis]  | 55    | 41      | 201         |
| 582       | 1      | 751        | 452       | [gi11256643]    | 20.2% identity with NADH dehydrogenase of the Leishmania major mitochondrion; putative [Bacillus subtilis]  | 55    | 36      | 300         |
| 645       | 5      | 2260       | 2057      | [gi1210824]     | fusion protein F [bovine respiratory syncytial virus] p[01481]VGNZBA (fusion glycoprotein precursor - bovine respiratory syncytial virus (strain A51908)) | 55    | 25      | 204         |
| 672       | 2      | 957        | 2216      | [gi11511333]    | M. jannaschii predicted coding region M1322 [Methanococcus jannaschii]  | 55    | 36      | 1260        |
| 730       | 1      | 955        | 479       | [gi1537007]     | ORF_3379 [Escherichia coli]   | 55    | 30      | 477         |
| 737       | 1      | 1859       | 945       | [gi1526963]     | CC Site No. 18166 [Escherichia coli]  | 55    | 30      | 915         |
| 742       | 2      | 228        | 572       | [gi1304160]     | product unknown [Bacillus subtilis]   | 55    | 38      | 345         |
| 817       | 2      | 1211       | 903       | [gi11136289]    | histidine kinase A [Dictyostelium discoideum]   | 55    | 29      | 309         |
| 819       | 1      | 582        | 355       | [gi1558073]     | polymorphic antigen [Plasmodium falciparum]   | 55    | 22      | 228         |
| 832       | 2      | 1152       | 724       | [gi140367]      | ORF3 [Clostridium acetobutylicum]   | 55    | 32      | 429         |
| 840       | 1      | 769        | 386       | [gi11205875]    | isoprenoid synthase 1 [Haemophilus influenzae]  | 55    | 39      | 384         |
| 1021      | 1      | 23         | 529       | [gi148563]      | beta-lactamase [Yersinia enterocolitica]  | 55    | 38      | 507         |
| 1026      | 1      | 60         | 335       | [gi147804]      | Opp C (AAI-301) [Salmonella typhimurium]  | 55    | 26      | 276         |
| 1525      | 1      | 1          | 282       | [gi11477533]    | serA [Staphylococcus aureus]  | 55    | 29      | 282         |
| 1814      | 2      | 224        | 985       | [gi11046078]    | M. genitalium predicted coding region MG369 [Mycoplasma genitalium]   | 55    | 38      | 762         |
| 3254      | 1      | 427        | 254       | [gi1413968]     | lpa-44d gene product [Bacillus subtilis]  | 55    | 30      | 174         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | nduf gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 3695      | 1      | 686        | 345       | gi 216773       | haloacetate dehalogenase H-1 [Moraxella sp.]   | 55    | 32      | 342         |
| 3721      | 1      | 1          | 312       | gi 42029        | ORF1 gene product [Escherichia coli]   | 55    | 31      | 312         |
| 3799      | 1      | 3          | 272       | gi 42029        | ORF1 gene product [Escherichia coli]   | 55    | 38      | 270         |
| 3889      | 1      | 22         | 423       | gi 1129145      | acetyl-CoA C-acyltransferase (Mangifera indica)  | 55    | 45      | 402         |
| 3916      | 1      | 2          | 385       | gi 529754       | apoC [Streptococcus pyogenes]  | 55    | 38      | 384         |
| 3945      | 1      | 4          | 198       | gi 476252       | phase 1 flagellin [Salmonella enterica]  | 55    | 36      | 195         |
| 4074      | 1      | 488        | 246       | gi 42029        | ORF1 gene product [Escherichia coli]   | 55    | 38      | 243         |
| 4184      | 1      | 2          | 343       | gi 1524267      | unknown [Mycobacterium tuberculosis]   | 55    | 28      | 342         |
| 4284      | 1      | 14         | 206       | gi 1100774      | ferredoxin-dependent glutamate synthase [Synecocystis sp.]   | 55    | 36      | 195         |
| 4457      | 2      | 644        | 378       | gi 180189       | cerebellar-degeneration-related antigen (COR34) [Homo sapiens] gi 182737<br>cerebellar degeneration-associated protein [Homo sapiens]<br>pir[A2970]A2970 cerebellar degeneration-related protein - human | 55    | 38      | 267         |
| 4514      | 1      | 2          | 244       | gi 216773       | haloacetate dehalogenase H-1 [Moraxella sp.]   | 55    | 32      | 243         |
| 4599      | 1      | 432        | 217       | gi 1129145      | acetyl-CoA C-acyltransferase [Mangifera indica]  | 55    | 42      | 216         |
| 4606      | 1      | 416        | 210       | gi 386120       | myosin alpha heavy chain (S2 subfragment) [rabbits, masseter, epitide<br>Partial, 234 aa]  | 55    | 27      | 207         |
| 5         | 8      | 5348       | 4932      | gi 1536069      | ORF YBL047c [Saccharomyces cerevisiae]   | 54    | 27      | 417         |
| 12        | 7      | 7166       | 6165      | gi 1205504      | homoserine acetyltransferase [Haemophilus influenzae]  | 54    | 30      | 1002        |
| 23        | 16     | 17086      | 15326     | gi 474192       | lucC gene product [Escherichia coli]   | 54    | 31      | 1761        |
| 35        | 1      | 2          | 979       | gi 48034        | small subunit of soluble hydrogenase (AA 1-384) [Synecococcus sp.]<br>ir[S06919]HQCSS soluble hydrogenase (EC 1.12.-.-) small chain -<br>nechococcus sp. (PCC 6716)                                      | 54    | 36      | 978         |
| 37        | 11     | 9437       | 8667      | gi 537207       | ORF f277 [Escherichia coli]  | 54    | 38      | 771         |
| 37        | 12     | 8165       | 8332      | gi 1160967      | palmitoyl-protein thioesterase [Homo sapiens]  | 54    | 37      | 168         |
| 46        | 15     | 13025      | 13804     | gi 438473       | protein is hydrophobic, with homology to E. coli Prow; putative Bacillus<br>subtilis   | 54    | 28      | 780         |
| 56        | 2      | 203        | 736       | gi 1256139      | YbbJ [Bacillus subtilis]   | 54    | 34      | 534         |
| 57        | 13     | 11117      | 10179     | gi 1151248      | inosine-uridine preferring nucleoside hydrolase [Crithidia fasciculata]  | 54    | 32      | 939         |
| 66        | 2      | 516        | 1133      | gi 1335781      | Cap [Drosophila melanogaster]  | 54    | 29      | 618         |
| 70        | 10     | 8116       | 8646      | gi 1399823      | PhoE [Rhizobium meliloti]  | 54    | 31      | 531         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession  | match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|------------------|--|-------|---------|-------------|
| 70        | 15     | 12556      | 11801     | [sp P02983 TCR_S | TETRACYCLINE RESISTANCE PROTEIN  | 54    | 29      | 756         |
| 87        | 5      | 4915       | 5706      | [gi 11064811     | [function unknown [Bacillus subtilis]                                      | 54    | 33      | 792         |
| 92        | 4      | 3005       | 2289      | [gi 1205366      | [oligopeptide transport ATP-binding protein [Haemophilus influenzae]       | 54    | 33      | 717         |
| 103       | 2      | 2596       | 1556      | [gi 710495       | [protein kinase [Bacillus brevis]  | 54    | 33      | 1041        |
| 105       | 2      | 3585       | 2095      | [gi 143727       | [putative [Bacillus subtilis]  | 54    | 30      | 1491        |
| 112       | 4      | 2317       | 2732      | [gi 153724       | [HslC [Streptococcus pneumoniae]   | 54    | 41      | 396         |
| 127       | 2      | 1720       | 2493      | [gi 144297       | [acetyl esterase (XynC) [Caldoceillum saccharolyticum] pir B37202 B37202   | 54    | 34      | 774         |
| 138       | 5      | 1600       | 3306      | [gi 142473       | [pyruvate oxidase [Escherichia coli]                                       | 54    | 36      | 1707        |
| 152       | 2      | 525        | 1172      | [gi 1377834      | [unknown [Bacillus subtilis]   | 54    | 23      | 648         |
| 161       | 9      | 4831       | 5469      | [gi 1903305      | [ORF73 [Bacillus subtilis]   | 54    | 28      | 639         |
| 161       | 13     | 6694       | 7251      | [gi 1511039      | [phosphate transport system regulatory protein [Methanococcus jannaschii]  | 54    | 32      | 558         |
| 164       | 6      | 3263       | 4543      | [gi 1204976      | [prolyl-tRNA synthetase [Haemophilus influenzae]                           | 54    | 34      | 1281        |
| 164       | 20     | 21602      | 22243     | [gi 143582       | [spoIIIEA protein [Bacillus subtilis]                                      | 54    | 32      | 642         |
| 171       | 6      | 5683       | 4250      | [gi 1436865      | [malA] gene products [Bacillus stearothermophilus] pir S43914 S43914       | 54    | 37      | 1434        |
| 206       | 18     | 18208      | 19720     | [gi 1240016      | [hypothetical protein 1 - Bacillus stearothermophilus                      | 54    | 38      | 513         |
| 218       | 2      | 1090       | 1905      | [gi 1467378      | [R09E10.3 [Caenorhabditis elegans]   | 54    | 26      | 816         |
| 220       | 1      | 1322       | 663       | [gi 1353761      | [unknown [Bacillus subtilis]   | 54    | 22      | 660         |
| 220       | 13     | 12655      | 13059     | [pir S00485 S004 | [myosin II heavy chain [Neogleria fowleri]                                 | 54    | 35      | 405         |
| 221       | 3      | 2030       | 3709      | [gi 1303813      | [gene 11-1 protein precursor - Plasmodium falciparum (fragments)           | 54    | 34      | 1680        |
| 272       | 7      | 5055       | 4219      | [gi 162984       | [Yqew [Bacillus subtilis]  | 54    | 33      | 837         |
| 316       | 7      | 4141       | 4701      | [gi 1682769      | [arylamine N-acetyltransferase (AA 1-290) [Gallus gallus] ir S06652 MYCHV3 | 54    | 31      | 561         |
| 316       | 10     | 6996       | 8742      | [gi 1413951      | [arylamine N-acetyltransferase (EC 2.3.1.5) (clone NAT-3) - chicken        | 54    | 28      | 1749        |
| 338       | 3      | 3377       | 2214      | [gi 1490328      | [mccE gene product [Escherichia coli]                                      | 54    | 28      | 1164        |
| 341       | 4      | 3201       | 3614      | [gi 171959       | [lpa-27d gene product [Bacillus subtilis]                                  | 54    | 25      | 414         |
|           |        |            |           |                  | [ORF F (unidentified)]   |       |         |             |
|           |        |            |           |                  | [myosin-like protein [Saccharomyces cerevisiae]                            |       |         |             |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 346       | 1      | 1820       | 912       | gi 396400       | similar to eukaryotic Na <sup>+</sup> /H <sup>+</sup> exchangers [Escherichia coli] sp P32702 YJCE_ECOLI HYPOTHETICAL 60.5 KD PROTEIN IN SOKR-ACS INTERGENIC REGION (0549) | 54    | 34      | 909         |
| 348       | 2      | 623        | 1351      | gi 537109       | ORF_1343a [Escherichia coli]   | 54    | 34      | 729         |
| 378       | 2      | 1007       | 1942      | sp P02983 TCR_S | TETRACYCLINE RESISTANCE PROTEIN  | 54    | 31      | 936         |
| 408       | 6      | 4351       | 5301      | gi 474190       | lucA gene product [Escherichia coli]   | 54    | 29      | 951         |
| 444       | 9      | 7934       | 8854      | gi 216267       | ORF2 [Bacillus megaterium]   | 54    | 32      | 921         |
| 463       | 2      | 2717       | 2229      | gi 304160       | product unknown [Bacillus subtilis]  | 54    | 50      | 489         |
| 502       | 2      | 1696       | 1133      | gi 1205015      | hypothetical protein (sp P10120) [Haemophilus influenzae]  | 54    | 38      | 564         |
| 505       | 6      | 6262       | 5357      | gi 1500558      | [2-hydroxyhepta-2,4-diene-1,7-dioate isomerase [Methanococcus jannaschii]]   | 54    | 41      | 906         |
| 550       | 1      | 2736       | 1522      | gi 40100        | rodC (tag) polypeptide (AA 1-746) [Bacillus subtilis] tr S06049 S06049 rodC protein - Bacillus subtilis p P1485 TAGP_BACSU TECHOIC ACID BIOSYNTHESIS PROTEIN F.            | 54    | 35      | 1215        |
| 551       | 5      | 3305       | 4279      | gi 950197       | unknown [Corynebacterium glutamicum]   | 54    | 34      | 975         |
| 558       | 2      | 1356       | 958       | gi 485090       | No definition line found [Caenorhabditis elegans]  | 54    | 32      | 399         |
| 580       | 1      | 91         | 936       | gi 231906       | [fused envelope glycoprotein precursor [Friend spleen focus-forming virus]]  | 54    | 45      | 846         |
| 603       | 3      | 554        | 757       | gi 1323423      | ORF_YGR234w [Saccharomyces cerevisiae]   | 54    | 36      | 204         |
| 617       | 1      | 25         | 249       | gi 219959       | [ornithine transcarbamylase [Homo sapiens]]  | 54    | 40      | 225         |
| 622       | 3      | 1097       | 1480      | gi 1303873      | Ygg2 [Bacillus subtilis]   | 54    | 25      | 384         |
| 623       | 1      | 3          | 404       | gi 1063250      | low homology to P20 protein of Bacillus licheniformis and bleomycin acetyltransferase of Streptomyces verticillus [Bacillus subtilis]                                      | 54    | 45      | 402         |
| 689       | 1      | 1547       | 1011      | gi 552446       | NADH dehydrogenase subunit 4 [Apis mellifera ligustica] pir S52968 S52968 NADH dehydrogenase chain 4 - honeybee mitochondrion (SGC4)                                       | 54    | 30      | 537         |
| 725       | 2      | 686        | 1441      | gi 987096       | [sensory protein kinase [Streptomyces hygroscopicus]]  | 54    | 26      | 756         |
| 956       | 1      | 1          | 249       | pir S30782 S307 | Integrin homolog - yeast [Saccharomyces cerevisiae]  | 54    | 24      | 249         |
| 978       | 2      | 1137       | 859       | gi 1130194      | ORF_YNL091w [Saccharomyces cerevisiae]   | 54    | 33      | 279         |
| 1314      | 1      | 3          | 381       | gi 1001108      | hypothetical protein [Synecocystis sp.]  | 54    | 33      | 279         |
| 2450      | 1      | 1          | 228       | gi 1045057      | ch-TGG [Homo sapiens]  | 54    | 32      | 228         |
| 2934      | 1      | 1          | 387       | gi 580870       | [ipa-37d qoxA gene produ. (Bacillus subtilis)]   | 54    | 36      | 387         |
| 2970      | 1      | 499        | 251       | sp P3734P YECE  | HYPOTHETICAL PROTEIN IN ASP5 5'-REGION (FRAGMENT)  | 54    | 42      | 249         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | na/cp gene name  | sim | ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-----|-------|-------------|
| 3002      | 1      | 1          | 309       | gi 44027        | [Tma protein [Lactococcus lactis]  | 54  | 33    | 309         |
| 3561      | 1      | 9          | 464       | gi 151259       | [HMG-CoA reductase (EC 1.1.1.88) (Pseudomonas mevalonii) pir A44756 A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.                                     | 54  | 35    | 456         |
| 3572      | 1      | 72         | 401       | gi 450688       | [hcdH gene of EcoRPII gene product [Escherichia coli] pir S38437 S38437 hcdH protein - Escherichia coli pir S09623 S09623 hypothetical protein A - Escherichia coli (SUN 40-520) | 54  | 36    | 330         |
| 3829      | 1      | 798        | 400       | gi 1322245      | [nevalonate pyrophosphate decarboxylase [Rattus norvegicus]  | 54  | 29    | 399         |
| 3909      | 1      | 1          | 273       | gi 29865        | [CENP-E (Homo sapiens)   | 54  | 30    | 273         |
| 3921      | 1      | 3          | 209       | pir S24325 S243 | [glucan 1,4-beta-glucosidase (EC 3.2.1.74) - Pseudomonas fluorescens subsp. cellulosa  | 54  | 34    | 207         |
| 4438      | 1      | 566        | 285       | gi 1196657      | [unknown protein [Mycoplasma pneumoniae]   | 54  | 30    | 282         |
| 4459      | 1      | 3          | 272       | gi 1046081      | [hypothetical protein (GB:026185.10) [Mycoplasma genitalium]   | 54  | 38    | 270         |
| 4564      | 1      | 3          | 221       | gi 216267       | [ORF2 [Bacillus megaterium]  | 54  | 38    | 219         |
| 23        | 12     | 12538      | 10685     | gi 474192       | [lucC gene product [Escherichia coli]  | 53  | 35    | 1854        |
| 23        | 14     | 14841      | 13579     | gi 42029        | [ORF1 gene product [Escherichia coli]  | 53  | 32    | 1263        |
| 24        | 3      | 4440       | 3940      | gi 1359947      | [c2 gene product [Bacteriophage B1]  | 53  | 36    | 501         |
| 26        | 4      | 3818       | 4618      | gi 1486247      | [unknown [Bacillus subtilis]   | 53  | 37    | 801         |
| 38        | 6      | 2856       | 3998      | gi 405880       | [yell [Escherichia coli]   | 53  | 40    | 1143        |
| 38        | 10     | 9380       | 7806      | gi 1399954      | [thyroid sodium/iodide symporter NIS [Rattus norvegicus]   | 53  | 29    | 1575        |
| 56        | 10     | 12324      | 12100     | pir A54592 A545 | [110K actin filament associated protein - chicken  | 53  | 32    | 225         |
| 57        | 6      | 5047       | 4583      | pir A00361 DEZP | [alcohol dehydrogenase (EC 1.1.1.1) - fission yeast (Schizosaccharomyces pombe)  | 53  | 39    | 465         |
| 57        | 12     | 10515      | 8932      | gi 480429       | [putative transcriptional regulator [Bacillus stearotherophilus]   | 53  | 30    | 1584        |
| 67        | 12     | 9496       | 10218     | gi 1511555      | [quinolone resistance norA protein protein [Methanococcus jannaschii]  | 53  | 31    | 723         |
| 69        | 3      | 3125       | 2382      | gi 1687017      | [arabinogalactan-protein, ADP [Nicotiana glauca, cell-suspension culture filtrate, Peptide, 461 aa]  | 53  | 30    | 744         |
| 79        | 1      | 3          | 1031      | gi 1523802      | [glucanase [Anabaena variabilis]   | 53  | 32    | 1029        |
| 80        | 1      | 673        | 338       | gi 452428       | [ATPase 3 [Plasmodium falciparum]  | 53  | 36    | 336         |
| 88        | 4      | 1910       | 2524      | gi 37034        | [ORF_0488 [Escherichia coli]   | 53  | 25    | 615         |
| 88        | 5      | 2467       | 3282      | gi 537034       | [ORF_0488 [Escherichia coli]   | 53  | 29    | 816         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match | Gene name  | Accession      | Match | Accession | Length (nt) |
|-----------|--------|------------|-----------|-------|--|----------------|-------|-----------|-------------|
| 92        | 8      | 5870       | 5505      | 91    | amphotropic murine retrovirus receptor [Rattus norvegicus]   | gi139598       | 53    | 33        | 366         |
| 94        | 5      | 4417       | 3239      | 91    | tropomyosin (TPM1) [Saccharomyces cerevisiae]  | gi173038       | 53    | 25        | 1179        |
| 99        | 5      | 4207       | 5433      | 91    | BICYCLOHEPTANESULFONAMIDE RESISTANCE PROTEIN (SULFONAMIDE RESISTANCE PROTEIN)  | spP28246 BCKLE | 53    | 30        | 1227        |
| 120       | 3      | 1639       | 2262      | 91    | ORF1 [Vibrio anguillarum]  | gi1576455      | 53    | 35        | 624         |
| 120       | 11     | 7257       | 8897      | 91    | glycine betaine transporter Opud [Bacillus subtilis]   | gi11324397     | 53    | 33        | 1641        |
| 127       | 6      | 6893       | 5685      | 91    | putative [Bacillus subtilis]   | gi11256630     | 53    | 32        | 1209        |
| 147       | 2      | 255        | 557       | 91    | epib gene product [Staphylococcus epidermidis]   | gi1581648      | 53    | 34        | 103         |
| 158       | 4      | 4705       | 4256      | 91    | mucoidy regulatory protein Algr [Pseudomonas aeruginosa] pIR[A32802 A32802 regulatory protein algr - Pseudomonas aeruginosa sp P26775 ALGR_PSEAE POSITIVE ALGINATE BIOSYNTHESIS REGULATORY PROTEIN]  | gi1151004      | 53    | 32        | 450         |
| 171       | 7      | 5717       | 5421      | 91    | hypothetical protein (GP.D6404.18) [Methanococcus jannaschii]  | gi11510669     | 53    | 34        | 297         |
| 191       | 9      | 13087      | 11483     | 91    | acetate decarboxylase [Clostridium acetobutylicum] pIR[B49346 B49346 butyrate-acetate CoA-transferase (EC 8.3.9) small chain - Clostridium acetobutylicum sp P13752 CTPA_CLOAB BUTYRATE-ACETONACETATE COA-TRANSFERASE SUBUNIT (EC 2.8.3.9) (COAT A)] | gi1298085      | 53    | 31        | 1805        |
| 203       | 5      | 3763       | 4326      | 91    | proE protein (ttg start codon) [Bacillus subtilis]   | gi143456       | 53    | 29        | 564         |
| 206       | 17     | 18204      | 16971     | 91    | acetylglutamate kinase [Bacillus stearothermophilus] sp Q07905 ARGB_BACST ACETYLGLUTAMATE KINASE (EC 2.7.2.8) (NAG INASE) (NAG) (N-ACETYL-L-GLUTAMATE 5-PHOSPHOTRANSFERASE)]   | gi1304136      | 53    | 36        | 768         |
| 212       | 10     | 4021       | 4221      | 91    | protein kinase [Plasmodium falciparum]   | gi19878        | 53    | 28        | 201         |
| 231       | 2      | 1580       | 1350      | 91    | paramyosin [Dirofilaria immitis]   | gi1537506      | 53    | 34        | 231         |
| 272       | 6      | 2719       | 3249      | 91    | hypothetical protein (gcid 3' region) - Streptococcus mutans   | gi133141 A331  | 53    | 34        | 531         |
| 308       | 3      | 927        | 2576      | 91    | ORF_0696 [Escherichia coli]  | gi1606252      | 53    | 33        | 1650        |
| 320       | 7      | 5645       | 5884      | 91    | RNA polymerase III largest subunit [Plasmodium falciparum] sp P27625 RPCL_PLAPA DNA-DIRECTED RNA POLYMERASE III LARGEST SUBUNIT (EC 2.7.7.6)   | gi1160596      | 53    | 33        | 240         |
| 327       | 1      | 218        | 901       | 91    | unknown [Schizosaccharomyces pombe]  | gi1854601      | 53    | 31        | 684         |
| 341       | 2      | 212        | 2500      | 91    | ORF1 [Campylobacter jejuni]  | gi1633732      | 53    | 31        | 2289        |
| 351       | 1      | 763        | 183       | 91    | HYPOTHETICAL 42.7 KD PROTEIN IN TSPA-LEUD INTERGENIC REGION (ORF104)   | sp P31675 YABM | 53    | 32        | 381         |
| 433       | 7      | 5087       | 4731      | 91    | MHC class II analog [Staphylococcus aureus]  | gi11001961     | 53    | 30        | 357         |
| 454       | 2      | 1240       | 980       | 91    | 40K cell wall protein precursor (ser 5' region) - Streptococcus mutans (strain OMZ175, serotype F)   | gi160328 A603  | 53    | 27        | 261         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | RefSeq gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 470       | 4      | 1123       | 1761      | gi 516826       | rat GCP360 (Rattus rattus)   | 53    | 30      | 639         |
| 483       | 1      | 432        | 217       | gi 1480429      | putative transcriptional regulator (Bacillus stearothermophilus)   | 53    | 33      | 216         |
| 544       | 1      | 516        | 1259      | gi 46587        | ORF 1 (AA 1 - 121) (1 is 2nd base in codon) (Staphylococcus aureus) ir S15765 hypothetical protein 1 (h1b 5' region) - aphyllococcus aureus (fragment)                         | 53    | 38      | 744         |
| 558       | 10     | 3957       | 3754      | gi 151140       | tes gene (Bacteriophage P1)  | 53    | 32      | 204         |
| 603       | 2      | 339        | 620       | gi 307738       | Hmp (Vibrio parahaemolyticus)  | 53    | 26      | 282         |
| 693       | 1      | 1669       | 941       | gi 153123       | toxic shock syndrome toxin-1 precursor (Staphylococcus aureus) p1r A24606/XCSA1 toxic shock syndrome toxin-1 precursor - Laphylococcus aureus                                  | 53    | 38      | 729         |
| 766       | 1      | 2          | 673       | gi 687600       | orfA2: orfA2 forms an operon with orfA1 (Listeria monocytogenes)   | 53    | 43      | 672         |
| 781       | 3      | 667        | 335       | gi 1204551      | pilin biogenesis protein (Haemophilus influenzae)  | 53    | 26      | 333         |
| 801       | 1      | 3          | 545       | gi 1279400      | SepA protein (Escherichia coli)  | 53    | 25      | 543         |
| 803       | 1      | 2          | 910       | gi 695278       | lipase-like enzyme (Alcaligenes eutrophus)   | 53    | 30      | 909         |
| 822       | 1      | 1177       | 590       | gi 298032       | EF (Streptococcus suis)  | 53    | 30      | 588         |
| 910       | 1      | 2          | 184       | gi 1044936      | unknown (Schizosaccharomyces pombe)  | 53    | 29      | 183         |
| 943       | 1      | 794        | 399       | gi 190508       | similar to unidentified ORF near 47 minutes (Escherichia coli) sp P1436 YICK_ECOLI HYPOTHETICAL 43.5 KD PROTEIN IN SELC-NLPA INTERGENIC REGION.                                | 53    | 30      | 396         |
| 988       | 1      | 1004       | 504       | gi 142441       | ORF 3: putative (Bacillus subtilis)  | 53    | 28      | 501         |
| 1064      | 1      | 3          | 434       | gi 305080       | myosin heavy chain (Entamoeba histolytica)   | 53    | 26      | 432         |
| 1366      | 1      | 3          | 452       | gi 308852       | transmembrane protein (Lactococcus lactis)   | 53    | 33      | 450         |
| 1758      | 1      | 792        | 397       | gi 1001774      | hypothetical protein (Synecococcus sp.)  | 53    | 30      | 396         |
| 1897      | 1      | 1          | 447       | gi 1303949      | YqjX (Bacillus subtilis)   | 53    | 27      | 447         |
| 2381      | 1      | 798        | 400       | gi 1146243      | 22.4% identity with Escherichia coli DNA-damage inducible protein ...; putative (Bacillus subtilis)  | 53    | 37      | 399         |
| 3537      | 1      | 1          | 327       | gi 450688       | hsdM gene of EcoPCR gene product (Escherichia coli) p1r S38437 S38437 hsdM protein - Escherichia coli p1r S09629 S09629 hypothetical protein A - Escherichia coli (SUB 40-520) | 53    | 35      | 327         |
| 3747      | 2      | 117        | 397       | gi 1477486      | transposase (Burkholderia cepacia)   | 53    | 53      | 261         |
| 31        | 5      | 3049       | 3441      | gi 868224       | Mo definition line found (Caenorhabditis elegans)  | 52    | 33      | 393         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 15        | 5      | 2205       | 2369      | gi 215966       | gal protein (gag start codon) [Bacteriophage 74]  | 52    | 34      | 165         |
| 19        | 3      | 2429       | 3808      | gi 1205379      | UDP-murac-pentapeptide synthetase [Haemophilus influenzae]  | 52    | 31      | 1380        |
| 24        | 1      | 6920       | 3462      | gi 579124       | predicted 86.4kd protein; 52kd observed [Mycobacteriophage 15]<br>pir S30971 S30971 gene 26 protein - Mycobacterium phage L5<br>sp Q05233 UG26.BPM1.5 MINOR TAIL PROTEIN OP26. (SUB 2-837)  | 52    | 32      | 3459        |
| 37        | 5      | 3015       | 3935      | gi 1500543      | PI15 protein [Methanococcus jannaschii]   | 52    | 25      | 921         |
| 38        | 13     | 8795       | 9703      | gi 46851        | glucose kinase [Streptomyces coelicolor]  | 52    | 29      | 909         |
| 44        | 16     | 10617      | 11066     | gi 42012        | peaE gene product [Escherichia coli]  | 52    | 36      | 450         |
| 46        | 1      | 3          | 521       | gi 1040957      | NADH dehydrogenase subunit 6 [Anopheles trinkae]  | 52    | 25      | 519         |
| 51        | 10     | 5531       | 6280      | gi 388269       | trac [Plasmodium falciparum]  | 52    | 32      | 750         |
| 56        | 5      | 3968       | 2826      | gi 181949       | endothelial differentiation protein (edg-1) [Homo sapiens]<br>pir A35300 A35300 G protein-coupled receptor edg-1 - human<br>sp P21453 EDG1_HUMAN PROBABLE G PROTEIN-COUPLED RECEPTOR EDG-1. | 52    | 23      | 1143        |
| 57        | 5      | 4850       | 4173      | gi 304153       | arbitol dehydrogenase [Bacillus subtilis]   | 52    | 27      | 678         |
| 62        | 5      | 3364       | 2870      | gi 1072399      | phaE gene product [Rhizobium meliloti]  | 52    | 25      | 495         |
| 62        | 6      | 4445       | 3651      | gi 46485        | NADH dehydrogenase [Synecococcus PCC7942]   | 52    | 27      | 795         |
| 67        | 14     | 11355      | 12962     | gi 1511365      | glutamate synthase (NADPH), subunit alpha [Methanococcus jannaschii]  | 52    | 30      | 1608        |
| 67        | 21     | 16935      | 18158     | gi 1204393      | hypothetical protein (SP:P1122) [Haemophilus influenzae]  | 52    | 25      | 1224        |
| 70        | 4      | 2185       | 1997      | gi 7227         | cytoplasmic dynein heavy chain [Dictyostelium discoideum] r A44357 A44357<br>dynein heavy chain, cytosolic - slime mold cytoskeleton [dictyostelium]  | 52    | 36      | 189         |
| 96        | 10     | 10005      | 10664     | gi 1408485      | B65G gene product [Bacillus subtilis]   | 52    | 26      | 660         |
| 103       | 5      | 3986       | 3351      | gi 1009368      | respiratory nitrate reductase [Bacillus subtilis]   | 52    | 42      | 636         |
| 109       | 3      | 4102       | 3350      | gi 699274       | lmbE gene product [Mycobacterium leprae]  | 52    | 39      | 751         |
| 109       | 119    | 15732      | 17300     | gi 1526981      | amino acid permease YeeF like protein [Salmonella typhimurium]  | 52    | 30      | 1569        |
| 121       | 3      | 1412       | 981       | gi 722911       | unknown [Saccharomyces cerevisiae]  | 52    | 32      | 432         |
| 125       | 3      | 865        | 1680      | gi 1294975      | put gene product [Porphyromonas gingivalis]   | 52    | 38      | 816         |
| 130       | 2      | 659        | 1807      | gi 1254634      | 25.8% identity over 120 aa with the Synecococcus sp. HpaV protein; putative [Bacillus subtilis]   | 52    | 36      | 1149        |
| 149       | 1      | 1164       | 583       | gi 1225943      | PBSX terminase [Bacillus subtilis]  | 52    | 33      | 582         |
| 149       | 14     | 4687       | 4415      | gi 1510368      | M. jannaschii predicted coding region M0272 [Methanococcus jannaschii]  | 52    | 35      | 273         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 167       | 1      | 216        | 1001      | gi1146025       | cell division protein [Escherichia coli]   | 52    | 43      | 786         |
| 188       | 1      | 120        | 1256      | gi11474915      | orf 337; translated orf similarity to SW: BCR_ECOLI bicyclomycin resistance protein of Escherichia coli [Coxiella burnetii] pir 544207 544207 hypothetical protein 337 - Coxiella burnetii (SUB-338)             | 52    | 26      | 1137        |
| 195       | 9      | 9161       | 8760      | gi11028         | mitochondrial outer membrane 72K protein (Neurospora crassa) r A36682 A36682 72K mitochondrial outer membrane protein - Neurospora crassa  | 52    | 25      | 402         |
| 200       | 3      | 2065       | 2607      | gi1142439       | ATP-dependent nuclease (Bacillus subtilis)   | 52    | 35      | 543         |
| 203       | 4      | 2776       | 3684      | gi11303698      | bltD (Bacillus subtilis)   | 52    | 25      | 909         |
| 227       | 8      | 5250       | 5651      | gi1305080       | myosin heavy chain (Entamoeba histolytica)   | 52    | 24      | 402         |
| 242       | 1      | 21         | 1424      | gi11060877      | Bary [Escherichia coli]  | 52    | 32      | 1604        |
| 249       | 5      | 4526       | 4753      | pir C3722 C372  | cytochrome P450 1A1, hepatic - dog (fragment)  | 52    | 23      | 228         |
| 255       | 1      | 2107       | 1055      | gi1143290       | penicillin-binding protein (Bacillus subtilis)   | 52    | 28      | 1053        |
| 276       | 7      | 3963       | 3664      | gi11001610      | hypothetical protein [Synecocystis sp.]  | 52    | 30      | 300         |
| 276       | 8      | 4456       | 4055      | gi1146235       | orf 13 [Mycoplasma capricolum]   | 52    | 26      | 402         |
| 289       | 2      | 1856       | 1449      | gi1150900       | GTP phosphohydrolase [Proteus vulgaris]  | 52    | 34      | 408         |
| 325       | 1      | 1          | 279       | gi11204874      | polypeptide deformylase (formylmethionine deformylase) [Haemophilus influenzae]  | 52    | 33      | 279         |
| 340       | 1      | 2017       | 1010      | gi11215695      | peptide transport system protein SapF homolog; SapF homolog [Mycoplasma pneumoniae]  | 52    | 33      | 1004        |
| 375       | 3      | 340        | 1878      | gi11467446      | similar to SpoVB [Bacillus subtilis]   | 52    | 28      | 1539        |
| 424       | 4      | 4104       | 3262      | gi11478239      | unknown [Mycobacterium tuberculosis]   | 52    | 34      | 843         |
| 430       | 1      | 3          | 575       | pir A42606 A426 | orfA 5' to orf405 - Saccharopolyspora erythraea (fragment)   | 52    | 28      | 573         |
| 444       | 4      | 4728       | 3712      | gi11408494      | homologous to penicillin acylase [Bacillus subtilis]   | 52    | 31      | 1017        |
| 465       | 1      | 1802       | 903       | gi1143331       | alkaline phosphatase regulatory protein (Bacillus subtilis) pir A27450 A27450 regulatory protein phoR - Bacillus subtilis ap P23545 PHO-R-BACSU ALKALINE PHOSPHATASE SYNTHESIS SENSOR PROTEIN PHO R (BC 2.7.3.-) | 52    | 36      | 900         |
| 469       | 5      | 4705       | 4169      | gi1755152       | highly hydrophobic integral membrane protein [Bacillus subtilis] ap P42953 TAGC-BACSU TEICHOIC ACID TRANSLOCATION PERMEASE PROTEIN AGG.  | 52    | 32      | 537         |
| 495       | 1      | 1262       | 633       | gi11204607      | transcription activator [Haemophilus influenzae]   | 52    | 25      | 630         |
| 505       | 7      | 6004       | 5762      | gi1142440       | ATP-dependent nuclease [Bacillus subtilis]   | 52    | 28      | 243         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | sim | ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-----|-------|-------------|
| 517       | 2      | 1162       | 1614      | gi166162        | Bacteriophage phi-11 int gene activator (Staphylococcus acteriophage phi 11)  | 52  | 35    | 453         |
| 543       | 2      | 444        | 1295      | gi1215693       | putative orf; GT9_0r434 [Mycoplasma pneumoniae]   | 52  | 25    | 852         |
| 586       | 1      | 1          | 336       | gi181648        | epib gene product (Staphylococcus epidermidis)  | 52  | 36    | 336         |
| 773       | 1      | 848        | 426       | gi1279769       | [FdhC (Methanobacterium thermoformicum)]  | 52  | 30    | 423         |
| 1120      | 2      | 100        | 330       | gi142439        | ATP-dependent nuclease [Bacillus subtilis]  | 52  | 35    | 231         |
| 1614      | 1      | 691        | 347       | gi289262        | [comE ORF] [Bacillus subtilis]  | 52  | 28    | 345         |
| 2495      | 1      | 1          | 324       | gi216151        | DNA polymerase (gene b; tgg start codon) [Bacteriophage SPO2] gi1579197<br>SP02 DNA polymerase (aa 3-648) [Bacteriophage SPO2] pirA21498 [BUBS2 DNA-directed DNA polymerase (EC 2.7.7.7) - phage P02] | 52  | 34    | 324         |
| 2931      | 1      | 566        | 285       | gi1256136       | Ybdc [Bacillus subtilis]  | 52  | 30    | 282         |
| 2943      | 1      | 577        | 320       | gi131713        | [HLA ORF (AA 1-245) [Escherichia coli]]   | 52  | 35    | 258         |
| 2993      | 1      | 588        | 295       | gi1298032       | [EF [Streptococcus suis]]   | 52  | 34    | 294         |
| 3667      | 1      | 612        | 307       | gi1849025       | hypothetical 64.7-kDa protein [Bacillus subtilis]   | 52  | 36    | 306         |
| 3944      | 1      | 478        | 280       | gi1218040       | [BAA [Bacillus licheniformis]]  | 52  | 36    | 219         |
| 3954      | 2      | 613        | 347       | gi1854064       | [U87 Human herpesvirus 6]   | 52  | 50    | 267         |
| 3986      | 1      | 90         | 401       | gi1205919       | [Na+ and Cl- dependent gamma-aminobutyric acid transporter [Haemophilus influenzae]]  | 52  | 33    | 312         |
| 4002      | 1      | 3          | 389       | gi140003        | [oxoglutarate dehydrogenase (HADP+)] [Bacillus subtilis] p123129 [ODO3_BACSU 2-OXOGLOUTARATE DEHYDROGENASE E1 COMPONENT (EC 2.4.2) (ALPHA- KETOGLUTARATE DEHYDROGENASE)]                              | 52  | 42    | 387         |
| 4020      | 1      | 1          | 249       | gi159388        | [ornithine decarboxylase [Leishmania donovani]]   | 52  | 47    | 249         |
| 4098      | 1      | 438        | 220       | gi1409795       | [No definition line found [Escherichia coli]]   | 52  | 32    | 219         |
| 4248      | 1      | 3          | 312       | gi1965077       | [Adp6 [Saccharomyces cerevisiae]]   | 52  | 40    | 210         |
| 7         | 1      | 3          | 575       | gi1895747       | [putative cel operon regulator [Bacillus subtilis]]   | 51  | 28    | 573         |
| 21        | 4      | 2479       | 3276      | gi1510962       | [indole-3-glycerol phosphate synthase [Methanococcus jannaschii]]   | 51  | 32    | 798         |
| 22        | 9      | 5301       | 5966      | gi1303933       | [yqjN [Bacillus subtilis]]  | 51  | 25    | 666         |
| 43        | 3      | 1516       | 1283      | gi1519460       | [Srp1 [Schizosaccharomyces pombe]]  | 51  | 31    | 234         |
| 44        | 17     | 11042      | 11305     | gi142011        | [moaD gene product [Escherichia coli]]  | 51  | 35    | 264         |
| 51        | 11     | 6453       | 6731      | gi1495471       | [vacuolating toxin [Helicobacter pylori]]   | 51  | 37    | 279         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 52        | 4      | 2537       | 2995      | gi1256652       | 25% identity to the E.coli regulatory protein MprA; putative [Bacillus subtilis]   | 51    | 32      | 459         |
| 57        | 10     | 7331       | 6843      | gi1508173       | RTA domain of PTS-dependent Gal transport and phosphorylation Escherichia coli   | 51    | 32      | 489         |
| 59        | 1      | 29         | 1111      | gi1299163       | alanine dehydrogenase [Bacillus subtilis]  | 51    | 33      | 1083        |
| 67        | 20     | 15791      | 16576     | gi1510977       | M. jannaschii predicted coding region M0938 [Methanococcus jannaschii]   | 51    | 24      | 786         |
| 69        | 2      | 1559       | 1218      | gi1467359       | unknown [Bacillus subtilis]  | 51    | 34      | 342         |
| 71        | 1      | 3          | 1196      | gi1298032       | EF [Streptococcus suis]  | 51    | 32      | 1194        |
| 78        | 2      | 149        | 176       | gi1161242       | proliferating cell nuclear antigen [Styela clavel]   | 51    | 28      | 174         |
| 99        | 4      | 3357       | 4040      | gi1642795       | TFIID subunit TAF1155 [Homo sapiens]   | 51    | 25      | 664         |
| 109       | 1      | 2852       | 1428      | gi1580920       | rodd (gtAA) polypeptide (AA 1-673) [Bacillus subtilis] pif[S06048]S06048 probable rodd protein - Bacillus subtilis sp133484[7AGE.BACSO PROBABLE POLY(GLYCEROL-PHOSPHATE) LPH-A-GLUCOSYLTRANSFERASE (EC 2.4.1.52) (TECHNOIC ACID BIOSYNTHESIS PROTEIN E)] | 51    | 27      | 1425        |
| 109       | 9      | 6007       | 6693      | gi11204815      | hypothetical protein (SP:P32662) [Haemophilus influenzae]  | 51    | 23      | 687         |
| 112       | 3      | 1066       | 2352      | pir[S05330]S053 | maltose-binding protein precursor - Enterobacter aerogenes   | 51    | 42      | 1287        |
| 112       | 13     | 14432      | 12855     | gi1405857       | yehU [Escherichia coli]  | 51    | 29      | 1578        |
| 114       | 9      | 9725       | 8967      | gi1435098       | orf1 [Mycoplasma capricolum]   | 51    | 30      | 759         |
| 115       | 1      | 1          | 912       | gi11431110      | ORF YHJ08Jw [Saccharomycus curvifilus]   | 51    | 25      | 912         |
| 127       | 10     | 9647       | 10477     | gi1204314       | H. influenzae predicted coding region H10056 [Haemophilus influenzae]  | 51    | 37      | 831         |
| 132       | 9      | 6814       | 7356      | gi1431929       | Muni regulatory protein [Mycoplasma sp.]   | 51    | 38      | 543         |
| 154       | 2      | 575        | 1153      | gi11237044      | unknown [Mycobacterium tuberculosis]   | 51    | 36      | 579         |
| 154       | 7      | 6587       | 5634      | gi1409286       | bcrU [Bacillus subtilis]   | 51    | 27      | 954         |
| 171       | 8      | 6943       | 6236      | gi11203484      | hypothetical protein (SP:P3918) [Haemophilus influenzae]   | 51    | 32      | 708         |
| 184       | 1      | 1          | 291       | gi1466886       | B1496_CJ_206 [Mycobacterium leprae]  | 51    | 33      | 291         |
| 212       | 5      | 1501       | 2139      | pir[A45605]A456 | mature-parasite-infected erythrocyte surface antigen MESA - Plasmodium falciparum  | 51    | 23      | 639         |
| 228       | 2      | 707        | 1378      | gi18204         | nuclear protein [Drosophila melanogaster]  | 51    | 27      | 672         |
| 236       | 8      | 8137       | 7481      | gi149272        | Asparaginase [Bacillus licheniformis]  | 51    | 31      | 657         |
| 243       | 4      | 4637       | 3546      | gi1151102       | malvalonate kinase [Methanococcus jannaschii]  | 51    | 29      | 1092        |

TABLE 2

S aureus - putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match<br>rcession | match gene name   | % sim | % ident | length<br>(nt) |
|-----------|--------|------------|-----------|-------------------|---|-------|---------|----------------|
| 257       | 4      | 3540       | 3373      | gi 204579         | H. influenzae predicted coding region H10326 [Haemophilus influenzae]   | 51    | 22      | 168            |
| 258       | 3      | 2397       | 1609      | gi 160299         | glutamic acid-rich protein [Plasmodium falciparum] pfr A54514 A54514<br>glutamic acid-rich protein precursor - Plasmodium falciparum                                    | 51    | 34      | 789            |
| 265       | 5      | 2419       | 3591      | gi 580841         | Fl [Bacillus subtilis]  | 51    | 32      | 1173           |
| 298       | 2      | 518        | 748       | gi 1336162        | SCP8 [Streptococcus agalactiae]   | 51    | 34      | 231            |
| 316       | 9      | 5817       | 7049      | gi 413953         | [pa-29d gene product [Bacillus subtilis]  | 51    | 39      | 1233           |
| 332       | 2      | 3775       | 2057      | gi 1208012        | mucS [Thermus aquaticus thermophilus]   | 51    | 26      | 1719           |
| 364       | 4      | 3816       | 4991      | gi 528991         | unknown [Bacillus subtilis]   | 51    | 32      | 1176           |
| 440       | 2      | 448        | 684       | gi 2819           | transferase (GAL10) (AA 1 - 687) [Kluyveromyces fragilis] r S01407 XUVRG<br>UDP-glucose 4-epimerase (EC 5.1.3.2) - yeast uvreromyces fragilis var. lactis               | 51    | 32      | 237            |
| 495       | 2      | 1353       | 1177      | gi 297861         | protease G [Erwinia chrysanthemi]   | 51    | 41      | 177            |
| 495       | 3      | 2287       | 1718      | gi 1513317        | serine rich protein [Entamoeba histolytica]   | 51    | 25      | 570            |
| 506       | 1      | 840        | 421       | gi 455320         | CII protein [Bacteriophage P4]  | 51    | 33      | 420            |
| 600       | 1      | 1474       | 983       | gi 587532         | orf. len: 201, CAI: 0.16 [Saccharomyces cerevisiae] pfr S48818 S48818<br>hypothetical protein yeast [Saccharomyces cerevisiae]  | 51    | 30      | 492            |
| 607       | 3      | 479        | 934       | gi 1511524        | hypothetical protein (SP-P37002) [Methanococcus jannaschii]   | 51    | 40      | 456            |
| 686       | 2      | 127        | 600       | gi 493017         | endocarditis specific antigen [Enterococcus faecalis]   | 51    | 30      | 474            |
| 726       | 1      | 33         | 230       | gi 1353851        | unknown [Prochlorococcus marinus]   | 51    | 45      | 198            |
| 861       | 1      | 176        | 652       | gi 410145         | dehydroquinase dehydratase [Bacillus subtilis]  | 51    | 34      | 477            |
| 869       | 1      | 782        | 393       | gi 40100          | rodC (tag) polypeptide (AA 1-746) [Bacillus subtilis] ir S06049 S06049<br>rodC protein - Bacillus subtilis p P13485 TAGC_BACSU TECHOTIC ACID<br>BIOSYNTHESIS PROTEIN F. | 51    | 23      | 390            |
| 1003      | 1      | 642        | 322       | gi 1279707        | hypothetical phosphoglycerate mutase [Saccharomyces cerevisiae]   | 51    | 39      | 321            |
| 1046      | 2      | 866        | 624       | gi 510257         | glycosyltransferase [Escherichia coli]  | 51    | 29      | 243            |
| 1467      | 1      | 702        | 352       | gi 1511175        | M. jannaschii predicted coding region MJ1177 [Methanococcus jannaschii]   | 51    | 32      | 351            |
| 2558      | 1      | 457        | 230       | sp P10582 DROM    | DNA POLYMERASE (EC 2.7.7.7) (S-1 DNA ORF 3).  | 51    | 26      | 228            |
| 3003      | 1      | 779        | 399       | gi 809543         | CbrC protein [Erwinia chrysanthemi]   | 51    | 27      | 381            |
| 3604      | 1      | 1          | 399       | gi JC4210 JC42    | 3-hydroxyacyl-CoA dehydrogenase (EC 1.1.1.35) - mouse   | 51    | 37      | 399            |
| 3732      | 1      | 2          | 316       | gi 145906         | acyl-CoA synthetase [Escherichia coli]  | 51    | 33      | 315            |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 3791      | 1      | 2          | 274       | gi1061351       | semaphorin III family homolog [Homo sapiens]  | 51    | 37      | 273         |
| 3995      | 1      | 46         | 336       | gi1216346       | surfactin synthetase [Bacillus subtilis]  | 51    | 38      | 291         |
| 4193      | 1      | 612        | 307       | gi142749        | ribosomal protein L12 (AA 1-179) [Escherichia coli] trIS04776[XXECP] peptide N-acetyltransferase (tm) (EC 2.3.1.1) - chetichia coli | 51    | 25      | 306         |
| 4539      | 1      | 167        | 185       | gi1408494       | homologous to penicillin acylase [Bacillus subtilis]  | 51    | 40      | 183         |
| 4562      | 1      | 442        | 239       | gi1458280       | coded for by C. elegans cDNA cm01e7; similar to hydroxymethylglutaryl-CoA synthase [Caenorhabditis elegans]                         | 51    | 35      | 204         |
| 1         | 4      | 3576       | 4859      | gi1559160       | GRAIL score: null; cap site and late promoter motifs present; putative [Autographa californica nuclear polyhedrosis virus]          | 50    | 44      | 1284        |
| 11        | 7      | 4044       | 5165      | gi1146207       | putative [Bacillus subtilis]  | 50    | 35      | 1122        |
| 11        | 113    | 10509      | 9496      | gi1208451       | hypothetical protein [Synecocystis sp.]   | 50    | 39      | 1014        |
| 19        | 1      | 2034       | 1018      | gi1413966       | ipe-42d gene product [Bacillus subtilis]  | 50    | 29      | 1017        |
| 20        | 111    | 8586       | 8407      | gi1321159       | ORF YOR103W [Saccharomyces cerevisiae]  | 50    | 28      | 180         |
| 24        | 5      | 5408       | 4824      | gi1496280       | structural protein [Bacteriophage Tuc2009]  | 50    | 29      | 585         |
| 34        | 4      | 1926       | 2759      | gi1303966       | HqjO [Bacillus subtilis]  | 50    | 36      | 834         |
| 38        | 30     | 122865     | 23440     | gi1072179       | Similar to dihydroflavonol-4-reductase (maize, petunia, tomato) [Caenorhabditis elegans]  | 50    | 32      | 576         |
| 47        | 2      | 1705       | 2976      | gi153015        | FemA protein [Staphylococcus aureus]  | 50    | 29      | 1272        |
| 56        | 113    | 15290      | 15841     | gi1806096       | ORF_1187; end overlaps end of o100 by 14 bases; start overlaps t174, ther starts possible [Escherichia coli]                        | 50    | 30      | 562         |
| 57        | 1      | 2135       | 1077      | gi1640922       | xyitol dehydrogenase (unidentified hemiascomycete)  | 50    | 29      | 1059        |
| 58        | 2      | 628        | 1761      | gi143725        | putative [Bacillus subtilis]  | 50    | 29      | 1134        |
| 88        | 6      | 4393       | 3884      | gi1072179       | Similar to dihydroflavonol-4-reductase (maize, petunia, tomato) [Caenorhabditis elegans]  | 50    | 32      | 510         |
| 89        | 5      | 3700       | 3356      | gi1276658       | ORF174 gene product [Porphyra purpurea]   | 50    | 25      | 345         |
| 141       | 1      | 3          | 239       | gi1476024       | carbamoyl phosphate synthetase II [Plasmodium falciparum]   | 50    | 33      | 237         |
| 151       | 1      | 186        | 626       | gi1403441       | unknown [Mycobacterium tuberculosis]  | 50    | 35      | 441         |
| 166       | 7      | 11065      | 9623      | gi1895747       | putative cel operon regulator [Bacillus subtilis]   | 50    | 32      | 1443        |
| 201       | 6      | 5284       | 5096      | gi160229        | Circumsporozoite protein [Plasmodium reichenowi]  | 50    | 42      | 189         |
| 206       | 22     | 30784      | 29555     | gi11052754      | LarP integral membrane protein [Lactococcus lactis]   | 50    | 24      | 1230        |

TABLE 2

S aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 211       | 4      | 1523       | 1927      | gi 410131       | ORFX7 [Bacillus subtilis]  | 50    | 29      | 405         |
| 214       | 4      | 2411       | 3395      | sp P37348 VECE  | HYPOTHETICAL PROTEIN IN ASP5 5'-REGION (FRAGMENT)  | 50    | 37      | 885         |
| 228       | 7      | 5068       | 4406      | gi 313580       | envelope protein (human immunodeficiency virus type 1) p1RIS3835/S3835 envelope protein - human immunodeficiency virus type 1 (fragment) (SUB 1-77)  | 50    | 35      | 663         |
| 272       | 2      | 3048       | 1723      | gi 1408485      | B65G gene product [Bacillus subtilis]  | 50    | 22      | 1326        |
| 273       | 2      | 1616       | 984       | gi 184186       | phosphoglycerate mutase [Saccharomyces cerevisiae]   | 50    | 28      | 613         |
| 328       | 2      | 2507       | 1605      | gi 148896       | lipoprotein [Haemophilus influenzae]   | 50    | 26      | 903         |
| 332       | 4      | 5469       | 3802      | gi 1526547      | DNA polymerase family X [Thermus aquaticus]  | 50    | 27      | 1668        |
| 342       | 5      | 1473       | 3931      | gi 456562       | C-box binding factor [Dictyostelium discoideum]  | 50    | 35      | 459         |
| 352       | 1      | 1478       | 741       | gi 288301       | ORF2 gene product [Bacillus megaterium]  | 50    | 29      | 738         |
| 404       | 7      | 5299       | 5523      | gi 111665       | ORF2136 [Marchantia polymorpha]  | 50    | 27      | 225         |
| 420       | 3      | 650        | 1825      | gi 757842       | UDP-sugar hydrolase [Bacterichia coli]   | 50    | 30      | 1176        |
| 464       | 1      | 1          | 591       | gi 487282       | Na <sup>+</sup> -ATPase subunit J [Enterococcus hirae]   | 50    | 29      | 591         |
| 472       | 2      | 1418       | 864       | gi 551875       | BglR [Lactococcus lactis]  | 50    | 23      | 555         |
| 520       | 1      | 23         | 541       | gi 567036       | CapE [Staphylococcus aureus]   | 50    | 27      | 519         |
| 529       | 1      | 6          | 410       | gi 1258652      | 25% identity to the E.coli regulatory protein MprA; putative [Bacillus subtilis]   | 50    | 34      | 405         |
| 534       | 5      | 7726       | 6059      | gi 255671       | Isolated as a weak suppressor of a mutant of the subunit AC10 of DNA dependent RNA polymerase I and III [Saccharomyces cerevisiae]   | 50    | 18      | 1668        |
| 647       | 1      | 2990       | 1497      | gi 405568       | TraI protein shares sequence similarity with a family of topoisomerases [Plasmid pSK11]  | 50    | 31      | 1494        |
| 664       | 3      | 1133       | 711       | gi 410007       | leukocidin F component [Staphylococcus aureus, MRSA No. 4, Peptide, 23 aa]   | 50    | 32      | 423         |
| 678       | 1      | 1          | 627       | gi 238032       | IEF [Streptococcus suis]   | 50    | 29      | 627         |
| 755       | 3      | 947        | 1171      | gi 150572       | cytochrome c1 precursor (EC 1.10.2.2) [Paracoccus denitrificans] gi 45465 cytochrome c1 (AA 1-450) [Paracoccus denitrificans] pir C29413 C29413 ubiquinol--cytochrome-c reductase (EC 1.10.2.2) cytochrome c1 precursor - Paracoccus denitrificans sp p13627 CY1 | 50    | 37      | 225         |
| 827       | 1      | 1363       | 683       | gi 142020       | heterocyst differentiation protein (Anabaena sp.)  | 50    | 21      | 681         |
| 892       | 1      | 3          | 752       | gi 48485        | B65G gene product [Bacillus subtilis]  | 50    | 27      | 750         |
| 910       | 2      | 438        | 887       | gi 1-04727      | tyrosine-specific transport protein [Haemophilus influenzae]   | 50    | 25      | 450         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Function ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name   | % sim | % Ident | length (nt) |
|-------------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
|             | 933    | 1          | 524       | gi1205451       | [cell division inhibitor (Haemophilus influenzae)]                                      | 50    | 32      | 237         |
|             | 973    | 1          | 424       | gi1886947       | [orf3 gene product (Saccharomyces cerevisiae)]  | 50    | 40      | 189         |
|             | 1009   | 1          | 653       | gi153727        | [M protein (group C streptococcus)]   | 50    | 28      | 225         |
|             | 1027   | 1          | 511       | gi1413934       | [ipa-10r gene product (Bacillus subtilis)]  | 50    | 25      | 255         |
|             | 1153   | 2          | 556       | gi1773676       | [ncca (Alcaligenes xylosoxydans)]   | 50    | 36      | 231         |
|             | 1222   | 1          | 798       | gi11408485      | [B65G gene product (Bacillus subtilis)]   | 50    | 21      | 399         |
|             | 1350   | 1          | 692       | gi1289272       | [ferrichrome-binding protein (Bacillus subtilis)]                                       | 50    | 32      | 294         |
|             | 2945   | 1          | 366       | gi1171704       | [hexaprenyl pyrophosphate synthetase (COQ1) (Saccharomyces erevisiae)]                  | 50    | 34      | 183         |
|             | 2968   | 2          | 1604      | gi1397526       | [clumping factor (Staphylococcus aureus)]   | 50    | 33      | 801         |
|             | 2998   | 2          | 657       | gi1495696       | [F54E7.3 gene product (Caenorhabditis elegans)]   | 50    | 40      | 264         |
|             | 3046   | 2          | 506       | ind1513819 5138 | [acyl carrier protein - Anabaena variabilis (fragment)]                                 | 50    | 32      | 201         |
|             | 3063   | 1          | 547       | gi1174190       | [lucA gene product (Escherichia coli)]  | 50    | 29      | 273         |
|             | 3174   | 1          | 3         | gi1151900       | [alcohol dehydrogenase (Rhodobacter sphaeroides)]                                       | 50    | 31      | 144         |
|             | 3792   | 1          | 625       | gi11001423      | [hypothetical protein (Synechocystis sp.)]  | 50    | 35      | 312         |
|             | 3800   | 1          | 2         | gi1144733       | [NAD-dependent beta-hydroxybutyryl coenzyme A dehydrogenase Clostridium acetobutylicum] | 50    | 24      | 261         |
|             | 3946   | 1          | 373       | gi1576765       | [cytochrome b (Myrmecia pilosula)]  | 50    | 38      | 146         |
|             | 3984   | 1          | 578       | gi1373481 ECE   | [HYPOTHETICAL PROTEIN IN ASP5 5' REGION (FRAGMENT)]                                     | 50    | 37      | 288         |
|             | 37     | 10         | 8250      | gi11204367      | [hypothetical protein (GB-U14003.278) (Haemophilus influenzae)]                         | 49    | 30      | 366         |
|             | 46     | 16         | 13802     | gi1466860       | [acd; B1308_F1_34 (Mycobacterium leprae)]   | 49    | 24      | 1047        |
|             | 59     | 5          | 2267      | gi1606304       | [ORF_0462 (Escherichia coli)]   | 49    | 27      | 1335        |
|             | 112    | 18         | 17884     | gi1559502       | [N04 protein (AA 1 - 409) (Caenorhabditis elegans)]                                     | 49    | 25      | 732         |
|             | 138    | 9          | 6973      | gi1303953       | [esterase (Acinetobacter calcoaceticus)]  | 49    | 29      | 930         |
|             | 217    | 6          | 4401      | gi1496234       | [fibrinectin/fibrinogen-binding protein (Streptococcus pyogenes)]                       | 49    | 31      | 738         |
|             | 220    | 12         | 11803     | gi1397526       | [clumping factor (Staphylococcus aureus)]   | 49    | 31      | 855         |
|             | 228    | 4          | 1842      | gi1523692 5236  | [hypothetical protein 9 - Plasmodium falciparum]  | 49    | 24      | 651         |
|             | 268    | 1          | 5016      | gi1143047       | [ORF8 (Bacillus subtilis)]  | 49    | 26      | 2403        |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 271       | 2      | 1164       | 1173      | gi11001257      | hypothetical protein [Synechocystis sp.]  | 49    | 38      | 210         |
| 300       | 3      | 4340       | 3180      | gi11510796      | hypothetical protein (CP-X91006.2) [Methanococcus jannaschii]   | 49    | 26      | 1161        |
| 381       | 1      | 2281       | 1142      | gi1396301       | Matches P50047: Bacterial regulatory proteins, arcC family signature [Escherichia coli]   | 49    | 29      | 1140        |
| 466       | 1      | 3          | 947       | gi11303863      | YggP [Bacillus subtilis]  | 49    | 26      | 945         |
| 666       | 1      | 379        | 191       | gi1633112       | ORF1 [Streptococcus sobrinus]   | 49    | 29      | 189         |
| 670       | 2      | 403        | 1014      | gi11122758      | unknown [Bacillus subtilis]   | 49    | 32      | 612         |
| 709       | 1      | 1433       | 795       | gi1143830       | XpaC [Bacillus subtilis]  | 49    | 29      | 639         |
| 831       | 1      | 943        | 473       | gi1401786       | phosphomannomutase [Mycoplasma pirum]   | 49    | 29      | 471         |
| 1052      | 1      | 422        | 213       | gi1303799       | YgeN [Bacillus subtilis]  | 49    | 21      | 210         |
| 1800      | 1      | 342        | 172       | gi1216300       | peptidoglycan synthesis enzyme [Bacillus subtilis] sp P37505 HURG_BACSU HURG PROTEIN UPD-N-ACETYLGLUCOSAMINE-N-ACETYLURAMYL-PENTAPEPTIDE PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE RANSEFRASE | 49    | 28      | 171         |
| 2430      | 1      | 2          | 376       | sp P27434 VFCA_ | HYPOTHETICAL 36.2 KD PROTEIN IN NDK-GCPE INTERGENIC REGION.   | 49    | 26      | 375         |
| 3096      | 1      | 542        | 273       | gi1516360       | surfactin synthetase [Bacillus subtilis]  | 49    | 25      | 270         |
| 32        | 4      | 3771       | 3100      | gi12117963      | hepatocyte nuclear factor 4 gamma [HNF4gamma] [Homo sapiens]  | 48    | 36      | 672         |
| 38        | 1      | 1          | 609       | gi1205790       | H. influenzae predicted coding region H11555 [Haemophilus influenzae]   | 48    | 28      | 609         |
| 45        | 6      | 5021       | 6427      | gi1534267       | unknown [Mycobacterium tuberculosis]  | 48    | 20      | 1407        |
| 59        | 14     | 16346      | 31096     | gi1197336       | LmpJ protein [Mycoplasma hominis]   | 48    | 28      | 14751       |
| 61        | 1      | 3          | 608       | gi1511555       | quinolone resistance nraA protein [Methanococcus jannaschii]  | 48    | 30      | 606         |
| 61        | 3      | 3311       | 3646      | gi1303893       | YqhL [Bacillus subtilis]  | 48    | 29      | 336         |
| 114       | 1      | 98         | 415       | gi1671708       | AutA homolog: similar to Drosophila melanogaster suppressor of able (AutA) protein, Swiss-Prot Accession Number P22193 [Drosophila virilis]   | 48    | 25      | 318         |
| 121       | 1      | 1131       | 610       | gi1314584       | unknown [Sphingomonas 588]  | 48    | 29      | 522         |
| 136       | 1      | 2014       | 1280      | gi1205968       | H. influenzae predicted coding region H11738 [Haemophilus influenzae]   | 48    | 23      | 735         |
| 171       | 10     | 8220       | 9557      | gi1208454       | hypothetical protein [Synechocystis sp.]  | 48    | 34      | 1338        |
| 175       | 1      | 3625       | 1814      | gi1396400       | similar to eukaryotic Na+/H+ exchanger [Escherichia coli] sp P32703 YJCE_ECOLI HYPOTHETICAL 60.5 KD PROTEIN IN SOXK-ACS INTERGENIC REGION [0549]  | 48    | 29      | 1812        |
| 194       | 1      | 2          | 385       | gi11510493      | H. jannaschii predicted coding region M20419 [Methanococcus jannaschii]   | 48    | 25      | 384         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Function ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name  | % sim | % ident | length (nt) |
|-------------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 197         | 1      | 901        | 452       | gi 1045714      | [spermidine/putrescine transport ATP-binding protein [Mycoplasma genitalium]   | 48    | 25      | 450         |
| 203         | 1      | 1          | 396       | gi 940288       | [protein localized in the nucleoli of pea nuclei; ORF; putative Pisum sativum]   | 48    | 29      | 396         |
| 204         | 1      | 1163       | 698       | gi 529202       | [No definition line found [Caenorhabditis elegans]   | 48    | 25      | 666         |
| 206         | 20     | 34815      | 27760     | gi 511490       | [gramicidin S synthetase 2 [Bacillus brevis]   | 48    | 27      | 7056        |
| 212         | 1      | 2          | 166       | gi 295899       | [nucleolin [Xenopus laevis]  | 48    | 34      | 165         |
| 220         | 10     | 12652      | 11426     | gi 44071        | [SecY protein [Lactococcus lactis]   | 48    | 23      | 1227        |
| 243         | 6      | 6450       | 5491      | gi 1184118      | [invasin kinase [Methanobacterium thermoautotrophicum]   | 48    | 30      | 960         |
| 264         | 4      | 5434       | 3308      | gi 1015903      | [ORF YJRI5c [Saccharomyces cerevisiae]   | 48    | 26      | 2127        |
| 441         | 1      | 1532       | 768       | gi 142863       | [replication initiation protein [Bacillus subtilis] pir B26580 B26580 replication initiation protein - Bacillus subtilis]    | 48    | 23      | 765         |
| 444         | 5      | 3898       | 5298      | gi 145836       | [putative [Escherichia coli]   | 48    | 24      | 1401        |
| 484         | 2      | 388        | 1110      | gi 146551       | [transmembrane protein (kdpD) [Escherichia coli]   | 48    | 18      | 723         |
| 542         | 3      | 1425       | 2000      | pir 528969 5289 | [N-carbamoylserine amidohydrolase (EC 3.5.1.59) - Arthrobacter sp.   | 48    | 27      | 576         |
| 566         | 1      | 1          | 1019      | gi 153490       | [tetracycline C resistance and export protein [Streptomyces laevis]  | 48    | 24      | 1017        |
| 611         | 1      | 2          | 730       | gi 1103507      | [unknown [Schizosaccharomyces pombe]   | 48    | 38      | 729         |
| 624         | 1      | 1255       | 665       | gi 144859       | [ORF B [Clostridium perfringens]   | 48    | 26      | 591         |
| 846         | 1      | 1014       | 508       | gi 537506       | [patamycin [Dirofilaria immitis]   | 48    | 27      | 507         |
| 1020        | 1      | 66         | 950       | gi 1499876      | [magnesium and cobalt transport protein [Methanococcus jannaschii]   | 48    | 30      | 885         |
| 1227        | 1      | 1          | 174       | gi 493730       | [lipoxigenase [Pisum sativum]  | 48    | 35      | 174         |
| 1266        | 1      | 1          | 405       | gi 882452       | [ORF (211); alternate name yggA; orf5 of X14436 [Escherichia coli] gi 41425 ORF5 (AA 1-197) [Escherichia coli] (SUB 15-211)] | 48    | 24      | 405         |
| 2071        | 1      | 707        | 381       | gi 1408486      | [HSTAA gene product [Bacillus subtilis]  | 48    | 25      | 327         |
| 2398        | 1      | 463        | 233       | gi 1500401      | [reverse gyrase [Methanococcus jannaschii]   | 48    | 40      | 231         |
| 2425        | 1      | 476        | 246       | pir H48563 H485 | [G1 protein - fowlpox virus (strain HP444) (fragment)]   | 48    | 40      | 231         |
| 2432        | 1      | 446        | 225       | gi 1353703      | [Trlo [Homo sapiens]   | 48    | 33      | 222         |
| 2453        | 1      | 794        | 399       | gi 142850       | [division initiation protein [Bacillus subtilis]   | 48    | 29      | 396         |
| 2998        | 1      | 469        | 236       | gi 577569       | [PepV [Lactobacillus delbrueckii]  | 48    | 31      | 234         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 3042      | 1      | 14         | 280       | gi 945219       | mucin (Homo sapiens)   | 48    | 35      | 267         |
| 3686      | 1      | 1          | 405       | gi 145816       | putative [Escherichia coli]  | 48    | 25      | 405         |
| 4027      | 2      | 492        | 301       | pir-SS117 S511  | trans-activator protein - Equine infectious anemia virus   | 48    | 32      | 192         |
| 4         | 2      | 3641       | 2232      | gi 1303989      | YqkI (Bacillus subtilis)   | 47    | 24      | 1410        |
| 24        | 2      | 599        | 1084      | gi 140083       | PC4-1 gene product (Bradyzia hygida)   | 47    | 28      | 486         |
| 36        | 10     | 7524       | 6925      | gi 1209223      | esterase (Acinetobacter lwoffii)   | 47    | 26      | 600         |
| 43        | 2      | 196        | 1884      | gi 1603455      | unknown [Mycobacterium tuberculosis]   | 47    | 27      | 1689        |
| 44        | 22     | 16116      | 15108     | gi 1511555      | quinolone resistance norA protein protein (Methanococcus jannaschii)   | 47    | 31      | 1011        |
| 69        | 7      | 7141       | 6710      | gi 438466       | Possible operon with orfG. Hydrophilic, no homologue in the database; putative [Bacillus subtilis]   | 47    | 29      | 432         |
| 81        | 4      | 5022       | 4279      | gi 466882       | bpsI; B1496_C2_189 (Mycobacterium leprae)  | 47    | 24      | 744         |
| 120       | 12     | 9135       | 8863      | gi 927340       | D9509.27p; CAI: 0.12 (Saccharomyces cerevisiae)  | 47    | 38      | 273         |
| 142       | 1      | 2022       | 1174      | gi 486143       | ORF YML094w (Saccharomyces cerevisiae)   | 47    | 32      | 849         |
| 168       | 1      | 2178       | 1093      | gi 117254       | Hypothetical EcsB protein [Bacillus subtilis]  | 47    | 29      | 1086        |
| 263       | 1      | 1884       | 943       | gi 142822       | O-alanine racemase cds [Bacillus subtilis]   | 47    | 34      | 942         |
| 279       | 1      | 1109       | 561       | gi 516608       | 2 predicted membrane helices. homology with B. subtilis man Orf3 Rowland et. al. unpublished Accession number M74183; approximately 1 minutes on updated Ruffd map; putative [Escherichia coli] SP P37355 YFBB_ECOLI_HYPOTHETICAL_26.7_KD PROTEIN IN MEM-MEM | 47    | 31      | 549         |
| 345       | 2      | 2620       | 1676      | gi 1204835      | hippuricase [Haemophilus influenzae]   | 47    | 28      | 945         |
| 389       | 2      | 152        | 400       | gi 456562       | C-box binding factor [Dictyostellum discoideum]  | 47    | 32      | 249         |
| 391       | 1      | 1          | 831       | gi 1420856      | myo-inositol transporter [Schizosaccharomyces pombe]   | 47    | 19      | 831         |
| 404       | 3      | 2072       | 2773      | gi 1255425      | C33C8.2 gene product (Caenorhabditis elegans)  | 47    | 17      | 702         |
| 529       | 5      | 2145       | 3107      | gi 1303973      | YqjV (Bacillus subtilis)   | 47    | 29      | 963         |
| 565       | 2      | 2321       | 1257      | gi 142824       | processing protease (Bacillus subtilis)  | 47    | 28      | 1065        |
| 654       | 1      | 962        | 483       | gi 243353       | ORF 5' of ECRF3 (herpesvirus saimiri HVS; host-squirrel monkey, eptide, 407 aa)  | 47    | 23      | 480         |
| 692       | 1      | 115        | 633       | gi 150756       | 40 kDa protein (plasmid pUHL)  | 47    | 25      | 519         |
| 765       | 1      | 1634       | 819       | gi 1256621      | 26.7% of identity in 165 aa to a Thermophilic bacterium hypothetical protein 6; putative [Bacillus subtilis]   | 47    | 28      | 816         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|---|-------|---------|-------------|
| 825       | 2      | 211        | 1023      | gi 397526       | clumping factor [Staphylococcus aureus]   | 47    | 32      | 813         |
| 914       | 1      | 1          | 615       | gi 558073       | polymorphic antigen [Plasmodium falciparum]   | 47    | 29      | 615         |
| 1076      | 1      | 1          | 753       | gi 1147557      | Aspartate aminotransferase [Bacillus circulans]   | 47    | 33      | 753         |
| 1351      | 1      | 793        | 398       | gi 755153       | ATP-binding protein [Bacillus subtilis]   | 47    | 20      | 396         |
| 4182      | 1      | 3          | 293       | gi 145836       | putative [Escherichia coli]   | 47    | 24      | 291         |
| 5         | 6      | 4708       | 4361      | gi 105080       | myosin heavy chain [Encarnosa histolytica]  | 46    | 30      | 348         |
| 11        | 4      | 2777       | 3058      | gi 601639       | Yel04Op [Saccharomyces cerevisiae]  | 46    | 28      | 282         |
| 46        | 11     | 10518      | 10300     | gi 1246901      | ATP-dependent DNA ligase [Candida albicans]   | 46    | 28      | 219         |
| 61        | 4      | 3941       | 7930      | gi 1298032      | EF [Streptococcus suis]   | 46    | 35      | 3990        |
| 132       | 4      | 5028       | 4093      | gi 1511057      | hypothetical protein SP: P43869 [Methanococcus jannaschii]  | 46    | 25      | 936         |
| 170       | 4      | 4719       | 3652      | gi 551910/5519  | G4 protein - Sauroleishmania tarentolae   | 46    | 26      | 1068        |
| 191       | 7      | 9343       | 8284      | gi 1041334      | F5405.7 [Caenorhabditis elegans]  | 46    | 25      | 1260        |
| 253       | 1      | 1          | 396       | gi 1204449      | dihydrolipamide acetyltransferase [Haemophilus influenzae]  | 46    | 35      | 396         |
| 264       | 3      | 437        | 973       | gi 180189       | cerebellar-degeneration-related antigen (CDR34) [Homo sapiens]  | 46    | 29      | 517         |
|           |        |            |           | pir A2970 A2970 | cerebellar degeneration-related protein - human   |       |         |             |
| 273       | 1      | 485        | 285       | gi 607573       | envelope glycoprotein C2v3 region [Human immunodeficiency virus type 1]                                 | 46    | 35      | 201         |
| 350       | 1      | 3          | 563       | gi 537052       | ORF_f286 [Escherichia coli]   | 46    | 35      | 561         |
| 384       | 1      | 2          | 862       | gi 121884       | [urea?] amidolyase [Haemophilus influenzae]   | 46    | 31      | 861         |
| 410       | 4      | 1876       | 2490      | gi 1110518      | proton antiporter efflux pump [Mycobacterium smegmatis]   | 46    | 24      | 615         |
| 432       | 1      | 2663       | 1455      | gi 1197634      | orf4; putative transporter; Method: conceptual translation supplied by author [Mycobacterium smegmatis] | 46    | 27      | 1209        |
| 458       | 1      | 2419       | 1211      | gi 15470        | portal protein [Bacteriophage SP11]   | 46    | 30      | 1209        |
| 517       | 5      | 2477       | 4192      | gi 1523412      | orf5 [Bacteriophage A2]   | 46    | 23      | 1716        |
| 540       | 3      | 1512       | 1285      | gi 215635       | pacA [Bacteriophage P1]   | 46    | 30      | 238         |
| 587       | 2      | 849        | 1242      | gi 537148       | ORF_181 [Escherichia coli]  | 46    | 29      | 594         |
| 1218      | 1      | 747        | 391       | gi 1205456      | single-stranded-DNA-specific exonuclease [Haemophilus influenzae]                                       | 46    | 30      | 357         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match      | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|------------|---|-------|---------|-------------|
| 3685      | 1      | 1          | 402       | gi 450688  | hsdM gene of E. coli gene product [Escherichia coli] pir S38437 S38437 hsdM protein - Escherichia coli pir S09629 S09629 hypothetical protein A - Escherichia coli [SUB 40-520] | 46    | 33      | 402         |
| 4176      | 1      | 673        | 338       | gi 351460  | PFM-C.1 gene product [Xenopus laevis]   | 46    | 31      | 336         |
| 37        | 7      | 4813       | 5922      | gi 606084  | ORF_1408 [Escherichia coli]   | 45    | 24      | 1110        |
| 38        | 16     | 11699      | 12004     | gi 452192  | protein tyrosine phosphatase (PTP-BAS, type 2) [Homo sapiens]   | 45    | 24      | 306         |
| 87        | 2      | 1748       | 2407      | gi 1064813 | homologous to sp:PHOR_BACSU [Bacillus subtilis]   | 45    | 23      | 660         |
| 103       | 12     | 14182      | 13385     | gi 1001307 | hypothetical protein [Synecocystis sp.]   | 45    | 22      | 798         |
| 112       | 14     | 14791      | 13811     | gi 1204389 | H. influenzae predicted coding region H10131 [Haemophilus influenzae]   | 45    | 23      | 981         |
| 145       | 4      | 4483       | 3461      | gi 220578  | open reading frame [Mus musculus]   | 45    | 20      | 1023        |
| 170       | 6      | 6329       | 4965      | gi 238657  | AppC-cytochrome d oxidase, subunit I homolog [Escherichia coli, K12, aptC, 514 aa]  | 45    | 27      | 1365        |
| 206       | 2      | 5230       | 4346      | gi 1222056 | aminotransferase [Haemophilus influenzae]   | 45    | 27      | 885         |
| 228       | 1      | 60         | 716       | gi 160299  | glutamic acid-rich protein [Plasmodium falciparum] pir A54514 A54514 glutamic acid-rich protein precursor - Plasmodium falciparum   | 45    | 23      | 657         |
| 288       | 1      | 2          | 1015      | gi 1255425 | IC108.2 gene product [Caenorhabditis elegans]   | 45    | 23      | 1014        |
| 313       | 3      | 4339       | 3128      | gi 581140  | NADH dehydrogenase [Escherichia coli]   | 45    | 30      | 1212        |
| 332       | 1      | 914        | 459       | gi 870966  | F47A4.2 [Caenorhabditis elegans]  | 45    | 20      | 456         |
| 344       | 1      | 3          | 221       | gi 171225  | kinesin-related protein [Saccharomyces cerevisiae]  | 45    | 26      | 219         |
| 441       | 2      | 1501       | 1073      | gi 142863  | replication initiation protein [Bacillus subtilis] pir B26580 B26580 replication initiation protein - Bacillus subtilis   | 45    | 27      | 429         |
| 672       | 1      | 2          | 982       | gi 1511334 | M. jannaschii predicted coding region MJ1323 [Methanococcus jannaschii]   | 45    | 22      | 981         |
| 763       | 3      | 1345       | 851       | gi 606180  | ORF_2310 [Escherichia coli]   | 45    | 24      | 495         |
| 886       | 3      | 379        | 846       | gi 726426  | similar to protein kinases and C. elegans proteins F37C12.8 and F37C12.5 [Caenorhabditis elegans]   | 45    | 30      | 468         |
| 948       | 1      | 3          | 473       | gi 156400  | myosin heavy chain (isozyme unc-54) [Caenorhabditis elegans] pir A9358 MKM myosin heavy chain B - Caenorhabditis elegans sp P02566 MYSB.CAEEL MYOSIN HEAVY CHAIN B (MHC B)      | 45    | 25      | 471         |
| 1158      | 1      | 2          | 376       | gi 141155  | transmission-blocking target antigen [Plasmodium falciparum]  | 45    | 35      | 375         |
| 2551      | 1      | 4          | 285       | gi 1276705 | ORF287 gene product [Porphyra purpurea]   | 45    | 28      | 282         |
| 3967      | 1      | 42         | 374       | gi 976025  | hrrA [Escherichia coli]   | 45    | 28      | 333         |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession                    | match gene name   | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|------------------------------------|---|-------|---------|-------------|
| 52        | 7      | 6931       | 5846      | gi 467378                          | unknown [Bacillus subtilis]   | 44    | 22      | 1086        |
| 118       | 8      | 6475       | 6849      | gi 171028                          | thioredoxin II [Saccharomyces cerevisiae]   | 44    | 28      | 375         |
| 221       | 5      | 7032       | 5617      | gi 53490                           | tetracycline C resistance and export protein [Streptomyces lauecensis]  | 44    | 21      | 1416        |
| 252       | 2      | 1331       | 1122      | gi 1204989                         | hypothetical protein (GB:U00022.9) [Haemophilus influenzae]   | 44    | 30      | 210         |
| 263       | 2      | 3265       | 2093      | gi 1136221                         | carboxypeptidase [Sulfolobus solfataricus]  | 44    | 26      | 1173        |
| 305       | 4      | 4963       | 3524      | gi 1298822                         | orf1 gene product [Lactobacillus helveticus]  | 44    | 31      | 1440        |
| 543       | 3      | 1315       | 1833      | gi 362250                          | low homology to P20 protein of Bacillus licheniformis and bleomycin acetyltransferase of Streptomyces verticillius [Bacillus subtilis]  | 44    | 24      | 519         |
| 544       | 4      | 3942       | 4892      | gi 951460                          | FM-C.1 gene product [Xenopus laevis]  | 44    | 32      | 951         |
| 792       | 1      | 1224       | 613       | gi 205680                          | high molecular weight neurofilament [Rattus norvegicus]   | 44    | 28      | 612         |
| 40        | 18     | 11303      | 11911     | gi 4511614                         | molybdopterin-guanine dinucleotide biosynthesis protein A [Methanococcus jannaschii]  | 43    | 27      | 609         |
| 59        | 8      | 3665       | 5128      | gi 151490                          | tetracycline C resistance and export protein [Streptomyces lauecensis]  | 43    | 21      | 1464        |
| 59        | 110    | 5536       | 7527      | gi 153022                          | lipase [Staphylococcus epidermidis]   | 43    | 22      | 1992        |
| 99        | 1      | 1346       | 681       | gi 1418051                         | unknown [Mycobacterium tuberculosis]  | 43    | 21      | 666         |
| 110       | 8      | 9402       | 12134     | gi 397526                          | clumping factor [Staphylococcus aureus]   | 43    | 21      | 2733        |
| 432       | 3      | 2782       | 2303      | gi 1605401605                      | sporozoite surface protein 2 - Plasmodium yoelii (fragment)   | 43    | 29      | 480         |
| 519       | 3      | 2547       | 3122      | sp 006530 DMSU_1.8.2.1 (FC) (FCSD) | SULFIDE DEHYDROGENASE (FLAVOCYTOCHROME C) FLAVOPROTEIN CHAIN PRECURSOR (EC 1.8.2.1) (FC) (FCSD)   | 43    | 23      | 576         |
| 4         | 13     | 12053      | 13321     | gi 295671                          | selected as a weak suppressor of a mutant of the subunit AC40 of DNA dependent RNA polymerase I and III [Saccharomyces cerevisiae]  | 42    | 18      | 1269        |
| 94        | 2      | 1768       | 1091      | gi 501027                          | ORF2 [Trypanosoma brucei]   | 42    | 31      | 678         |
| 127       | 4      | 5791       | 4550      | gi 42029                           | ORF1 gene product [Escherichia coli]  | 42    | 21      | 1242        |
| 297       | 3      | 1515       | 1036      | gi 142790                          | ORF1; putative [Bacillus firmus]  | 42    | 25      | 480         |
| 144       | 6      | 4097       | 3525      | gi 40320                           | ORF 2 (AA 1-203) [Bacillus thuringiensis]   | 42    | 30      | 573         |
| 512       | 1      | 2167       | 1115      | gi 405957                          | yeef [Escherichia coli]   | 42    | 23      | 1053        |
| 631       | 1      | 2434       | 1223      | gi 580920                          | rodd (gtaA) polypeptide (AA 1-673) [Bacillus subtilis] pir 506048 506048 probable rodd protein - Bacillus subtilis sp P13484 TAGE_BACSU PROBABLE POLY (GLYCEROL-PHOSPHATE) LPHAS-GLUCOSYLTRANSFERASE (EC 2.4.1.52) (TECHOIC ACID BIOSYNTHESIS ROPEIN E) | 42    | 24      | 1212        |

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | match accession | match gene name  | % sim | % ident | length (nt) |
|-----------|--------|------------|-----------|-----------------|--|-------|---------|-------------|
| 685       | 3      | 2359       | 1719      | gi 1103784      | [QeD [Bacillus subtilis]   | 42    | 19      | 621         |
| 4132      | 1      | 787        | 395       | gi 1022910      | protein tyrosine phosphatase [Dictyostelium discoideum]  | 42    | 25      | 393         |
| 86        | 2      | 1375       | 884       | gi 109506       | spermidine/spermine N1-acetyltransferase [Mus saxicola] pir S31430 S31430 spermidine/spermine N1-acetyltransferase - spiny ouse [Mus saxicola]                                     | 41    | 30      | 492         |
| 191       | 12     | 14797      | 14075     | gi 1124957      | orf4 gene product [Methanosarcina barkeri]   | 41    | 22      | 723         |
| 212       | 6      | 2150       | 3127      | gi 115873       | observed 35.2kd protein [Mycobacteriophage 15]   | 41    | 26      | 978         |
| 213       | 3      | 1261       | 2000      | gi 633692       | TrsA [Yersinia enterocolitica]   | 41    | 18      | 738         |
| 408       | 4      | 2625       | 3386      | gi 1197634      | orf4; putative transporter; Method: conceptual translation supplied by author [Mycobacterium smegmatis]  | 41    | 24      | 762         |
| 542       | 1      | 3          | 1103      | gi 457146       | rhostry protein [Plasmodium yoelii]  | 41    | 21      | 1101        |
| 924       | 1      | 2          | 475       | pir JH0148 JH01 | nucleolin - rat  | 41    | 30      | 474         |
| 1562      | 1      | 1          | 402       | gi 552184       | asparagine-rich antigen PfA35-2 [Plasmodium falciparum] pir S27826 S27826 asparagine-rich antigen PfA35-2 - Plasmodium falciparum (fragment)                                       | 40    | 20      | 402         |
| 2395      | 1      | 518        | 261       | pir S42251 S422 | hypothetical protein 5 - fowlpox virus   | 40    | 18      | 258         |
| 4077      | 1      | 3          | 305       | gi 1055055      | coded for by C. elegans cDNA YK3791.5; coded for by C. elegans cDNA YK509.5; coded for by C. elegans cDNA YK189.5; alternatively spliced form of F52C9.8b [Caenorhabditis elegans] | 39    | 21      | 303         |
| 918       | 1      | 1003       | 503       | gi 1255425      | C3308.2 gene product [Caenorhabditis elegans]  | 37    | 25      | 501         |
| 59        | 12     | 8294       | 10636     | gi 535260       | STARP antigen [Plasmodium reichenowi]  | 36    | 24      | 2343        |
| 63        | 5      | 3550       | 8079      | gi 298032       | EF [Streptococcus suis]  | 36    | 19      | 4530        |
| 544       | 3      | 2507       | 3601      | gi 1015903      | ORF YJR151C [Saccharomyces cerevisiae]   | 35    | 22      | 1095        |
| 63        | 4      | 1949       | 3574      | gi 552195       | circumsporozoite protein [Plasmodium falciparum] sp P05691 CSP_PLAFL CIRCUMSPOROZOITE PROTEIN (CS) [FRAGMENT]  | 32    | 27      | 1626        |

TABLE 2

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|-----------|--------|------------|-----------|
| 4         | 1      | 1234       | 692       |
| 4         | 3      | 1712       | 2278      |
| 4         | 4      | 3703       | 3032      |
| 4         | 14     | 11073      | 11585     |
| 5         | 2      | 2539       | 1601      |
| 5         | 3      | 1532       | 1771      |
| 5         | 7      | 4741       | 4550      |
| 5         | 9      | 7839       | 8422      |
| 5         | 12     | 8711       | 8547      |
| 6         | 4      | 2359       | 1982      |
| 8         | 1      | 349        | 176       |
| 11        | 6      | 5144       | 5983      |
| 11        | 9      | 5968       | 6498      |
| 11        | 10     | 6472       | 6284      |
| 11        | 16     | 10954      | 11271     |
| 12        | 5      | 5352       | 4942      |
| 12        | 6      | 4596       | 4862      |
| 15        | 3      | 1895       | 1650      |
| 16        | 10     | 11263      | 10835     |
| 18        | 2      | 1093       | 917       |
| 20        | 9      | 9135       | 7764      |
| 20        | 10     | 8571       | 8230      |
| 20        | 12     | 9201       | 8803      |
| 20        | 13     | 11236      | 10470     |
| 21        | 1      | 674        | 339       |
| 23        | 6      | 6138       | 5485      |
| 23        | 8      | 6376       | 5942      |

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

| Coding<br>ID | ORF<br>ID | Start<br>(nt) | Stop<br>(nt) |
|--------------|-----------|---------------|--------------|
| 23           | 9         | 7651          | 6881         |
| 23           | 15        | 12618         | 12830        |
| 24           | 4         | 4556          | 4185         |
| 24           | 6         | 5642          | 5241         |
| 25           | 2         | 1824          | 2402         |
| 31           | 2         | 505           | 849          |
| 31           | 3         | 1177          | 1524         |
| 31           | 4         | 2454          | 3005         |
| 32           | 2         | 765           | 1388         |
| 32           | 9         | 7952          | 8575         |
| 32           | 10        | 8591          | 8728         |
| 32           | 11        | 9738          | 9779         |
| 32           | 12        | 10797         | 10087        |
| 34           | 2         | 1315          | 1049         |
| 36           | 7         | 5226          | 5801         |
| 36           | 11        | 3575          | 7261         |
| 36           | 12        | 7424          | 7621         |
| 37           | 4         | 3158          | 2964         |
| 38           | 2         | 1585          | 940          |
| 38           | 11        | 6425          | 6868         |
| 38           | 20        | 116982        | 116371       |
| 38           | 26        | 20253         | 20804        |
| 38           | 27        | 20722         | 21264        |
| 39           | 1         | 1             | 627          |
| 40           | 1         | 805           | 404          |
| 43           | 1         | 796           | 428          |
| 44           | 4         | 2674          | 2324         |

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|-----------|--------|------------|-----------|
| 44        | 5      | 2484       | 3263      |
| 44        | 14     | 10587      | 10129     |
| 44        | 20     | 13724      | 13536     |
| 44        | 21     | 13596      | 13994     |
| 45        | 7      | 6575       | 6297      |
| 46        | 8      | 6365       | 6520      |
| 46        | 12     | 10449      | 10976     |
| 46        | 17     | 15032      | 15424     |
| 47        | 1      | 288        | 1079      |
| 48        | 9      | 7620       | 7778      |
| 50        | 1      | 1612       | 962       |
| 50        | 2      | 1621       | 3316      |
| 51        | 1      | 738        | 370       |
| 51        | 5      | 2520       | 2245      |
| 53        | 1      | 442        | 287       |
| 53        | 7      | 6705       | 6319      |
| 54        | 7      | 9014       | 8709      |
| 55        | 1      | 592        | 326       |
| 55        | 3      | 1052       | 786       |
| 56        | 1      | 1          | 261       |
| 56        | 3      | 1551       | 1228      |
| 56        | 4      | 1970       | 1560      |
| 56        | 17     | 19092      | 18712     |
| 57        | 4      | 3694       | 3521      |
| 57        | 8      | 5436       | 5822      |
| 58        | 9      | 8885       | 8553      |
| 59        | 3      | 1366       | 1509      |

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

| Contig<br>ID | ORF<br>ID | Start<br>(nt) | Stop<br>(nt) |
|--------------|-----------|---------------|--------------|
| 59           | 6         | 3026          | 2802         |
| 59           | 7         | 3770          | 3570         |
| 59           | 9         | 4946          | 4583         |
| 59           | 11        | 7518          | 8378         |
| 59           | 13        | 110403        | 16403        |
| 62           | 2         | 2696          | 1521         |
| 62           | 11        | 5440          | 5757         |
| 63           | 1         | 1             | 336          |
| 67           | 1         | 900           | 1781         |
| 67           | 2         | 1774          | 2610         |
| 67           | 3         | 2591          | 3904         |
| 67           | 8         | 7110          | 6955         |
| 68           | 1         | 78            | 326          |
| 70           | 6         | 6761          | 5199         |
| 70           | 11        | 8335          | 8645         |
| 77           | 3         | 1590          | 1192         |
| 79           | 2         | 1509          | 1228         |
| 79           | 3         | 1411          | 1791         |
| 83           | 1         | 2             | 403          |
| 85           | 9         | 8300          | 8653         |
| 85           | 10        | 8969          | 8781         |
| 86           | 3         | 1426          | 1232         |
| 87           | 8         | 9187          | 9386         |
| 88           | 3         | 1820          | 1922         |
| 89           | 1         | 3             | 161          |
| 89           | 7         | 5042          | 4878         |
| 91           | 1         | 1098          | 550          |

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|-----------|--------|------------|-----------|
| 91        | 3      | 3938       | 3141      |
| 92        | 2      | 449        | 528       |
| 92        | 3      | 1958       | 1457      |
| 92        | 9      | 5638       | 6024      |
| 94        | 1      | 661        | 332       |
| 94        | 3      | 2445       | 1813      |
| 94        | 4      | 2583       | 2197      |
| 96        | 11     | 10601      | 11050     |
| 99        | 6      | 4672       | 4523      |
| 99        | 7      | 5014       | 4784      |
| 100       | 8      | 7658       | 7287      |
| 102       | 7      | 4697       | 4368      |
| 103       | 3      | 2496       | 2035      |
| 104       | 1      | 2          | 694       |
| 104       | 2      | 699        | 1277      |
| 105       | 1      | 1235       | 693       |
| 105       | 3      | 3233       | 2655      |
| 105       | 1      | 3          | 221       |
| 106       | 3      | 1209       | 1355      |
| 107       | 1      | 1081       | 542       |
| 109       | 4      | 4025       | 3651      |
| 109       | 13     | 11625      | 11996     |
| 109       | 14     | 11981      | 12268     |
| 109       | 20     | 17401      | 17688     |
| 110       | 1      | 2          | 760       |
| 114       | 10     | 8784       | 9384      |
| 116       | 1      | 1          | 309       |

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

| Contig No | ORF ID | Start (nt) | Stop (nt) |
|-----------|--------|------------|-----------|
| 116       | 3      | 6273       | 4462      |
| 116       | 8      | 11049      | 9976      |
| 116       | 9      | 10313      | 10158     |
| 120       | 5      | 3703       | 3320      |
| 120       | 6      | 4270       | 3869      |
| 120       | 13     | 9290       | 9844      |
| 121       | 2      | 417        | 569       |
| 126       | 3      | 1080       | 818       |
| 127       | 3      | 2648       | 3196      |
| 127       | 5      | 4084       | 4395      |
| 131       | 6      | 6773       | 6438      |
| 132       | 2      | 715        | 1695      |
| 134       | 1      | 2          | 667       |
| 135       | 2      | 512        | 258       |
| 135       | 3      | 1124       | 729       |
| 138       | 1      | 3          | 152       |
| 138       | 7      | 6008       | 4463      |
| 140       | 1      | 2660       | 1032      |
| 140       | 2      | 2019       | 1513      |
| 140       | 5      | 2387       | 2743      |
| 142       | 2      | 1360       | 2388      |
| 142       | 7      | 8830       | 7586      |
| 143       | 7      | 7290       | 6502      |
| 144       | 1      | 1227       | 640       |
| 146       | 1      | 2          | 511       |
| 146       | 3      | 502        | 1350      |
| 146       | 4      | 3673       | 2540      |

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|-----------|--------|------------|-----------|
| 146       | 5      | 2874       | 3071      |
| 147       | 1      | 1          | 339       |
| 149       | 11     | 3956       | 3615      |
| 149       | 12     | 4036       | 3785      |
| 149       | 13     | 4507       | 4145      |
| 149       | 15     | 4807       | 4610      |
| 149       | 16     | 5495       | 5049      |
| 149       | 18     | 5739       | 5491      |
| 149       | 21     | 7416       | 7054      |
| 149       | 23     | 9216       | 8521      |
| 149       | 24     | 9681       | 9106      |
| 149       | 25     | 10679      | 9897      |
| 150       | 2      | 2303       | 1587      |
| 154       | 3      | 1795       | 1508      |
| 154       | 8      | 6586       | 6398      |
| 154       | 14     | 12704      | 12147     |
| 154       | 15     | 13531      | 12803     |
| 156       | 1      | 315        | 593       |
| 157       | 3      | 1183       | 2232      |
| 158       | 2      | 1471       | 1064      |
| 159       | 3      | 452        | 808       |
| 161       | 2      | 876        | 1808      |
| 161       | 6      | 4653       | 4279      |
| 161       | 7      | 4803       | 4540      |
| 161       | 8      | 4896       | 4717      |
| 161       | 11     | 5817       | 5638      |
| 163       | 2      | 1604       | 840       |

TABLE 3

*S. aureus* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|-----------|--------|------------|-----------|
| 163       | 5      | 2796       | 2344      |
| 163       | 7      | 2952       | 2847      |
| 163       | 9      | 4905       | 5132      |
| 164       | 3      | 1338       | 1147      |
| 166       | 3      | 5213       | 4854      |
| 168       | 4      | 2500       | 2868      |
| 168       | 5      | 3595       | 4158      |
| 170       | 3      | 2517       | 2777      |
| 171       | 2      | 2277       | 1450      |
| 171       | 11     | 12576      | 11125     |
| 172       | 1      | 3          | 278       |
| 172       | 2      | 1940       | 1149      |
| 173       | 1      | 1289       | 708       |
| 173       | 5      | 7001       | 6114      |
| 174       | 2      | 593        | 1105      |
| 175       | 3      | 2552       | 2890      |
| 175       | 5      | 3820       | 3335      |
| 175       | 7      | 4342       | 4506      |
| 182       | 4      | 5477       | 4986      |
| 184       | 5      | 6043       | 5702      |
| 188       | 2      | 1210       | 1755      |
| 188       | 4      | 2647       | 2994      |
| 189       | 6      | 2614       | 3039      |
| 190       | 3      | 1998       | 2564      |
| 191       | 1      | 1          | 153       |
| 191       | 2      | 950        | 669       |
| 191       | 10     | 11786      | 13039     |

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|-----------|--------|------------|-----------|
| 191       | 11     | 12902      | 12163     |
| 192       | 1      | 91         | 426       |
| 195       | 3      | 2306       | 1932      |
| 195       | 5      | 2899       | 2606      |
| 198       | 2      | 1016       | 1591      |
| 201       | 1      | 170        | 625       |
| 203       | 2      | 783        | 1466      |
| 206       | 6      | 8910       | 7815      |
| 206       | 12     | 113947     | 113636    |
| 206       | 21     | 26208      | 27960     |
| 212       | 2      | 170        | 817       |
| 212       | 3      | 796        | 1167      |
| 212       | 7      | 3128       | 3436      |
| 212       | 9      | 3749       | 4075      |
| 213       | 1      | 1          | 705       |
| 214       | 2      | 1076       | 570       |
| 214       | 6      | 4064       | 3738      |
| 214       | 9      | 8600       | 6995      |
| 214       | 10     | 7864       | 7469      |
| 217       | 1      | 1927       | 965       |
| 218       | 1      | 178        | 657       |
| 218       | 3      | 1776       | 2156      |
| 220       | 2      | 1851       | 1369      |
| 220       | 3      | 1251       | 2262      |
| 220       | 7      | 8275       | 7208      |
| 220       | 8      | 10244      | 8661      |
| 220       | 9      | 11796      | 10216     |

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|-----------|--------|------------|-----------|
| 221       | 4      | 3095       | 2613      |
| 221       | 9      | 11428      | 10737     |
| 226       | 1      | 3          | 659       |
| 226       | 2      | 2196       | 1459      |
| 226       | 3      | 1476       | 1961      |
| 227       | 1      | 2          | 487       |
| 227       | 2      | 460        | 575       |
| 227       | 4      | 1855       | 2121      |
| 227       | 5      | 2052       | 2345      |
| 227       | 6      | 4760       | 3768      |
| 227       | 9      | 5591       | 6167      |
| 228       | 5      | 2503       | 2877      |
| 228       | 6      | 2846       | 3526      |
| 233       | 7      | 3944       | 3762      |
| 236       | 2      | 809        | 579       |
| 238       | 2      | 1975       | 1391      |
| 239       | 2      | 1417       | 905       |
| 241       | 5      | 4495       | 4334      |
| 242       | 2      | 1677       | 1363      |
| 243       | 1      | 127        | 576       |
| 244       | 1      | 1291       | 647       |
| 244       | 2      | 3035       | 1962      |
| 245       | 2      | 1614       | 1258      |
| 246       | 1      | 69         | 215       |
| 246       | 4      | 738        | 1733      |
| 249       | 3      | 3906       | 3712      |
| 250       | 1      | 494        | 249       |

S. aureus - putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|-----------|--------|------------|-----------|
| 254       | 1      | 1          | 156       |
| 256       | 2      | 956        | 1144      |
| 257       | 3      | 3700       | 3227      |
| 260       | 4      | 4906       | 4580      |
| 261       | 4      | 2196       | 2406      |
| 261       | 6      | 3214       | 3481      |
| 264       | 2      | 155        | 439       |
| 264       | 5      | 5252       | 4533      |
| 264       | 6      | 4739       | 5107      |
| 267       | 2      | 1323       | 931       |
| 268       | 4      | 5140       | 4700      |
| 272       | 1      | 862        | 446       |
| 272       | 3      | 1200       | 1439      |
| 272       | 9      | 4691       | 4909      |
| 272       | 10     | 6469       | 6035      |
| 276       | 4      | 1746       | 1901      |
| 278       | 1      | 224        | 553       |
| 278       | 5      | 3299       | 3448      |
| 278       | 7      | 4849       | 5127      |
| 285       | 2      | 551        | 736       |
| 288       | 3      | 1756       | 1950      |
| 288       | 5      | 2055       | 2276      |
| 289       | 1      | 2107       | 1055      |
| 290       | 2      | 2234       | 1932      |
| 291       | 2      | 332        | 622       |
| 291       | 5      | 1545       | 2051      |
| 295       | 3      | 1606       | 1349      |

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|-----------|--------|------------|-----------|
| 295       | 4      | 2728       | 2141      |
| 295       | 5      | 2220       | 2762      |
| 297       | 2      | 788        | 465       |
| 298       | 1      | 2          | 205       |
| 300       | 2      | 2380       | 1928      |
| 301       | 7      | 2794       | 2624      |
| 304       | 1      | 3          | 194       |
| 306       | 1      | 109        | 654       |
| 306       | 5      | 4016       | 4257      |
| 307       | 1      | 674        | 339       |
| 307       | 8      | 3645       | 3995      |
| 308       | 1      | 1          | 654       |
| 308       | 2      | 1120       | 599       |
| 308       | 4      | 2643       | 2332      |
| 313       | 2      | 2314       | 1919      |
| 314       | 1      | 10         | 702       |
| 316       | 2      | 982        | 1341      |
| 316       | 6      | 2758       | 3165      |
| 317       | 1      | 2          | 1114      |
| 317       | 3      | 4570       | 3458      |
| 321       | 6      | 5645       | 5217      |
| 321       | 7      | 6319       | 6140      |
| 321       | 8      | 7450       | 6794      |
| 322       | 2      | 827        | 543       |
| 326       | 2      | 165        | 1112      |
| 326       | 3      | 1117       | 1467      |
| 328       | 1      | 936        | 469       |

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

| Contig<br>ID | ORF<br>ID | Start<br>(nt) | Stop<br>(nt) |
|--------------|-----------|---------------|--------------|
| 328          | 5         | 3452          | 3276         |
| 329          | 1         | 3             | 719          |
| 329          | 2         | 781           | 1212         |
| 329          | 3         | 1471          | 1833         |
| 330          | 1         | 576           | 289          |
| 330          | 3         | 1447          | 1623         |
| 332          | 3         | 2353          | 2204         |
| 332          | 7         | 4971          | 5138         |
| 333          | 2         | 3295          | 3128         |
| 335          | 1         | 864           | 433          |
| 337          | 2         | 95            | 526          |
| 340          | 2         | 1658          | 1356         |
| 341          | 1         | 3             | 281          |
| 341          | 3         | 2476          | 3192         |
| 341          | 5         | 3618          | 3944         |
| 341          | 6         | 3929          | 4556         |
| 344          | 5         | 3197          | 2889         |
| 345          | 1         | 1532          | 768          |
| 346          | 2         | 221           | 592          |
| 350          | 3         | 1410          | 1598         |
| 352          | 2         | 2178          | 1765         |
| 352          | 3         | 7316          | 4596         |
| 352          | 7         | 7867          | 8404         |
| 352          | 8         | 8906          | 9247         |
| 352          | 9         | 10171         | 9854         |
| 359          | 1         | 1             | 546          |
| 362          | 1         | 3             | 656          |

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|-----------|--------|------------|-----------|
| 364       | 7      | 2158       | 1808      |
| 364       | 8      | 10574      | 10714     |
| 365       | 2      | 1612       | 1313      |
| 365       | 5      | 4680       | 4090      |
| 365       | 7      | 4980       | 6239      |
| 366       | 3      | 520        | 1719      |
| 367       | 3      | 906        | 1085      |
| 368       | 1      | 748        | 494       |
| 375       | 1      | 2          | 136       |
| 380       | 3      | 1351       | 1097      |
| 389       | 1      | 1          | 276       |
| 390       | 1      | 2          | 877       |
| 390       | 2      | 1373       | 1549      |
| 391       | 2      | 251        | 560       |
| 395       | 1      | 391        | 197       |
| 396       | 1      | 2132       | 1068      |
| 398       | 3      | 1344       | 1141      |
| 399       | 1      | 176        | 669       |
| 401       | 3      | 566        | 847       |
| 402       | 2      | 100        | 465       |
| 404       | 8      | 5561       | 5370      |
| 408       | 2      | 1507       | 2269      |
| 408       | 3      | 2875       | 2672      |
| 408       | 5      | 3524       | 4423      |
| 410       | 3      | 2111       | 1890      |
| 413       | 1      | 880        | 488       |
| 416       | 1      | 607        | 320       |

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|-----------|--------|------------|-----------|
| 416       | 2      | 578        | 847       |
| 416       | 3      | 2195       | 3590      |
| 417       | 1      | 3          | 179       |
| 417       | 2      | 161        | 616       |
| 420       | 2      | 788        | 513       |
| 422       | 2      | 357        | 677       |
| 431       | 2      | 856        | 1407      |
| 432       | 2      | 446        | 1084      |
| 433       | 1      | 1          | 417       |
| 433       | 3      | 2311       | 2033      |
| 434       | 1      | 942        | 535       |
| 434       | 2      | 2089       | 1235      |
| 440       | 1      | 1          | 450       |
| 442       | 2      | 1269       | 3320      |
| 443       | 3      | 1873       | 1520      |
| 444       | 1      | 1          | 696       |
| 444       | 7      | 6761       | 6366      |
| 451       | 1      | 940        | 614       |
| 453       | 2      | 896        | 636       |
| 453       | 8      | 3833       | 4786      |
| 453       | 9      | 4718       | 4512      |
| 453       | 10     | 4937       | 4731      |
| 455       | 1      | 434        | 219       |
| 455       | 2      | 472        | 930       |
| 459       | 1      | 265        | 687       |
| 462       | 1      | 2          | 247       |
| 466       | 2      | 1494       | 907       |

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

| Contig<br>ID | ORF<br>ID | Start<br>(nt) | Stop<br>(nt) |
|--------------|-----------|---------------|--------------|
| 467          | 1         | 654           | 349          |
| 468          | 1         | 2             | 250          |
| 469          | 1         | 1488          | 925          |
| 469          | 3         | 2386          | 3372         |
| 469          | 4         | 3464          | 3706         |
| 470          | 1         | 77            | 538          |
| 470          | 6         | 4098          | 3694         |
| 470          | 7         | 6330          | 5686         |
| 470          | 9         | 7351          | 8181         |
| 470          | 10        | 8175          | 9773         |
| 471          | 1         | 940           | 500          |
| 471          | 2         | 1562          | 1017         |
| 476          | 1         | 70            | 267          |
| 477          | 1         | 2             | 760          |
| 477          | 3         | 1764          | 2081         |
| 477          | 4         | 2066          | 2332         |
| 480          | 5         | 4016          | 4261         |
| 481          | 2         | 956           | 480          |
| 486          | 3         | 613           | 774          |
| 487          | 6         | 1795          | 2112         |
| 488          | 1         | 715           | 359          |
| 492          | 1         | 127           | 675          |
| 493          | 1         | 2             | 520          |
| 493          | 2         | 496           | 1242         |
| 502          | 3         | 1149          | 1571         |
| 504          | 1         | 690           | 346          |
| 505          | 5         | 4566          | 4150         |

S. aureus - Putative ending regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|-----------|--------|------------|-----------|
| 511       | 2      | 1741       | 1232      |
| 512       | 2      | 583        | 747       |
| 515       | 1      | 609        | 812       |
| 517       | 4      | 2179       | 2511      |
| 520       | 4      | 2897       | 2360      |
| 520       | 6      | 3908       | 3669      |
| 527       | 1      | 1          | 498       |
| 528       | 1      | 637        | 335       |
| 529       | 2      | 1679       | 1104      |
| 530       | 7      | 5298       | 5534      |
| 536       | 1      | 308        | 156       |
| 538       | 1      | 1362       | 736       |
| 538       | 3      | 2203       | 2880      |
| 538       | 5      | 3531       | 3121      |
| 538       | 6      | 4348       | 3731      |
| 540       | 1      | 996        | 664       |
| 540       | 2      | 1495       | 1031      |
| 541       | 1      | 89         | 433       |
| 541       | 2      | 719        | 432       |
| 542       | 2      | 1048       | 1272      |
| 545       | 2      | 1012       | 734       |
| 551       | 1      | 2145       | 1129      |
| 555       | 2      | 892        | 704       |
| 558       | 3      | 1357       | 1154      |
| 558       | 4      | 1760       | 1458      |
| 558       | 5      | 2105       | 1821      |
| 558       | 6      | 2166       | 2020      |

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

| Contig<br>ID | ORF<br>ID | Start<br>(nt) | Stop<br>(nt) |
|--------------|-----------|---------------|--------------|
| 558          | 7         | 2636          | 2322         |
| 558          | 8         | 3053          | 2802         |
| 558          | 9         | 3986          | 3453         |
| 560          | 1         | 475           | 921          |
| 565          | 3         | 1706          | 1485         |
| 571          | 1         | 308           | 156          |
| 571          | 3         | 994           | 1206         |
| 577          | 1         | 2             | 199          |
| 577          | 2         | 163           | 453          |
| 579          | 1         | 1             | 477          |
| 579          | 2         | 1784          | 1200         |
| 583          | 1         | 1988          | 996          |
| 585          | 1         | 946           | 539          |
| 587          | 1         | 22            | 573          |
| 588          | 2         | 1896          | 1372         |
| 588          | 3         | 1742          | 1554         |
| 590          | 1         | 47            | 334          |
| 592          | 2         | 1455          | 1141         |
| 593          | 1         | 2             | 775          |
| 593          | 2         | 817           | 1122         |
| 595          | 1         | 87            | 890          |
| 596          | 3         | 3593          | 1435         |
| 602          | 1         | 8             | 169          |
| 603          | 5         | 1071          | 1469         |
| 606          | 1         | 322           | 768          |
| 607          | 5         | 1444          | 1226         |
| 610          | 1         | 1029          | 541          |

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|-----------|--------|------------|-----------|
| 612       | 1      | 3          | 500       |
| 616       | 1      | 991        | 650       |
| 617       | 2      | 736        | 491       |
| 622       | 1      | 36         | 347       |
| 625       | 4      | 2046       | 2549      |
| 627       | 1      | 67         | 210       |
| 628       | 1      | 901        | 452       |
| 631       | 3      | 4789       | 4004      |
| 634       | 1      | 1448       | 759       |
| 636       | 1      | 189        | 368       |
| 636       | 2      | 1929       | 1063      |
| 637       | 2      | 2323       | 1954      |
| 638       | 1      | 227        | 1081      |
| 639       | 1      | 518        | 261       |
| 639       | 2      | 1377       | 811       |
| 641       | 1      | 118        | 444       |
| 642       | 3      | 1615       | 1331      |
| 642       | 4      | 2260       | 1847      |
| 643       | 1      | 3          | 608       |
| 645       | 4      | 1514       | 1758      |
| 645       | 6      | 2025       | 2321      |
| 645       | 7      | 2940       | 2488      |
| 648       | 1      | 2          | 1045      |
| 650       | 1      | 77         | 601       |
| 660       | 2      | 576        | 872       |
| 661       | 1      | 1725       | 961       |
| 664       | 2      | 89         | 304       |

TABLE 3

5. aureus - putative coding regions of novel proteins not similar to known proteins

| Contig<br>ID | ORF<br>ID | Start<br>(nt) | Stop<br>(nt) |
|--------------|-----------|---------------|--------------|
| 667          | 1         | 3             | 413          |
| 668          | 1         | 1             | 330          |
| 671          | 2         | 812           | 516          |
| 673          | 1         | 3             | 338          |
| 674          | 2         | 865           | 584          |
| 679          | 1         | 1             | 237          |
| 679          | 3         | 1589          | 1906         |
| 688          | 1         | 1236          | 835          |
| 688          | 2         | 1352          | 1077         |
| 694          | 1         | 3             | 143          |
| 696          | 2         | 818           | 432          |
| 706          | 1         | 36            | 224          |
| 709          | 3         | 1183          | 1449         |
| 711          | 1         | 3             | 908          |
| 715          | 1         | 3             | 167          |
| 716          | 1         | 2             | 637          |
| 721          | 1         | 133           | 570          |
| 722          | 1         | 763           | 383          |
| 723          | 1         | 1656          | 829          |
| 723          | 2         | 1498          | 1112         |
| 727          | 1         | 2             | 472          |
| 729          | 1         | 268           | 441          |
| 731          | 1         | 130           | 828          |
| 735          | 1         | 2             | 214          |
| 736          | 1         | 3             | 782          |
| 738          | 1         | 2             | 298          |
| 742          | 1         | 3             | 230          |

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|-----------|--------|------------|-----------|
| 745       | 3      | 1148       | 780       |
| 748       | 2      | 282        | 464       |
| 749       | 1      | 685        | 344       |
| 751       | 1      | 901        | 452       |
| 755       | 1      | 97         | 522       |
| 755       | 2      | 520        | 918       |
| 756       | 2      | 663        | 400       |
| 764       | 2      | 1033       | 746       |
| 767       | 1      | 1          | 405       |
| 768       | 1      | 2          | 373       |
| 771       | 1      | 1058       | 534       |
| 778       | 1      | 1735       | 902       |
| 785       | 1      | 1790       | 1023      |
| 787       | 1      | 1260       | 631       |
| 791       | 1      | 3          | 224       |
| 799       | 1      | 15         | 260       |
| 804       | 1      | 304        | 711       |
| 805       | 1      | 3          | 680       |
| 808       | 1      | 219        | 842       |
| 810       | 1      | 2221       | 1312      |
| 810       | 2      | 1774       | 1442      |
| 812       | 1      | 38         | 979       |
| 817       | 1      | 714        | 358       |
| 818       | 2      | 487        | 1104      |
| 819       | 2      | 1529       | 1032      |
| 819       | 3      | 1748       | 1419      |
| 820       | 1      | 195        | 1064      |

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

| Config | ORF | Start | Stop |
|--------|-----|-------|------|
| ID     | ID  | (nt)  | (nt) |
| 828    | 1   | 506   | 255  |
| 829    | 1   | 48    | 800  |
| 830    | 1   | 578   | 291  |
| 832    | 1   | 594   | 298  |
| 835    | 1   | 320   | 796  |
| 840    | 3   | 491   | 709  |
| 845    | 1   | 912   | 457  |
| 850    | 2   | 303   | 449  |
| 853    | 1   | 715   | 359  |
| 860    | 1   | 2     | 256  |
| 864    | 1   | 18    | 410  |
| 864    | 2   | 383   | 715  |
| 864    | 6   | 1676  | 1828 |
| 870    | 1   | 1     | 588  |
| 873    | 1   | 906   | 454  |
| 875    | 1   | 584   | 274  |
| 877    | 1   | 1661  | 1020 |
| 878    | 1   | 981   | 544  |
| 879    | 1   | 1567  | 785  |
| 881    | 1   | 1     | 243  |
| 882    | 1   | 389   | 604  |
| 890    | 1   | 2     | 508  |
| 905    | 1   | 793   | 398  |
| 906    | 1   | 852   | 544  |
| 912    | 1   | 373   | 188  |
| 913    | 1   | 3     | 290  |
| 913    | 2   | 1092  | 547  |

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|-----------|--------|------------|-----------|
| 915       | 1      | 6          | 161       |
| 915       | 2      | 169        | 402       |
| 921       | 1      | 126        | 386       |
| 927       | 1      | 1578       | 808       |
| 928       | 1      | 2          | 385       |
| 929       | 1      | 2          | 400       |
| 932       | 1      | 2          | 400       |
| 934       | 1      | 1          | 384       |
| 936       | 1      | 1052       | 528       |
| 937       | 1      | 2          | 616       |
| 945       | 1      | 220        | 645       |
| 945       | 2      | 649        | 1242      |
| 946       | 1      | 1702       | 950       |
| 949       | 1      | 1          | 270       |
| 951       | 1      | 3          | 362       |
| 955       | 1      | 3          | 143       |
| 960       | 1      | 723        | 400       |
| 963       | 1      | 1          | 162       |
| 965       | 1      | 590        | 346       |
| 966       | 1      | 1079       | 506       |
| 969       | 1      | 3          | 302       |
| 971       | 1      | 12         | 170       |
| 974       | 1      | 319        | 161       |
| 976       | 1      | 592        | 348       |
| 977       | 1      | 2          | 211       |
| 982       | 1      | 1926       | 982       |
| 984       | 1      | 589        | 296       |

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|-----------|--------|------------|-----------|
| 987       | 1      | 3          | 467       |
| 993       | 1      | 1          | 525       |
| 994       | 1      | 920        | 549       |
| 1004      | 1      | 557        | 318       |
| 1014      | 1      | 624        | 313       |
| 1015      | 1      | 2          | 463       |
| 1016      | 1      | 288        | 145       |
| 1019      | 1      | 1205       | 660       |
| 1022      | 1      | 839        | 474       |
| 1024      | 1      | 595        | 299       |
| 1024      | 2      | 276        | 431       |
| 1030      | 1      | 673        | 338       |
| 1032      | 1      | 355        | 179       |
| 1040      | 1      | 794        | 399       |
| 1043      | 1      | 3          | 269       |
| 1044      | 2      | 115        | 399       |
| 1047      | 1      | 1          | 159       |
| 1051      | 1      | 704        | 354       |
| 1051      | 2      | 1233       | 733       |
| 1063      | 1      | 2          | 400       |
| 1069      | 1      | 2          | 146       |
| 1069      | 2      | 769        | 533       |
| 1075      | 1      | 707        | 399       |
| 1077      | 1      | 97         | 405       |
| 1081      | 1      | 58         | 438       |
| 1086      | 1      | 1          | 384       |
| 1087      | 2      | 246        | 431       |

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

| Config<br>ID | ORF<br>ID | Start<br>(nt) | Stop<br>(nt) |
|--------------|-----------|---------------|--------------|
|              | 1088      | 1             | 374          |
|              | 1096      | 1             | 474          |
|              | 1098      | 1             | 1015         |
|              | 1100      | 1             | 1020         |
|              | 1100      | 2             | 1520         |
|              | 1101      | 1             | 703          |
|              | 1102      | 1             | 385          |
|              | 1107      | 1             | 2            |
|              | 1114      | 1             | 3            |
|              | 1115      | 1             | 2            |
|              | 1119      | 1             | 22           |
|              | 1129      | 1             | 40           |
|              | 1132      | 1             | 360          |
|              | 1133      | 1             | 609          |
|              | 1144      | 1             | 446          |
|              | 1147      | 1             | 528          |
|              | 1153      | 1             | 1            |
|              | 1154      | 1             | 3            |
|              | 1159      | 1             | 1            |
|              | 1161      | 1             | 341          |
|              | 1164      | 1             | 427          |
|              | 1171      | 1             | 19           |
|              | 1171      | 2             | 108          |
|              | 1183      | 1             | 2            |
|              | 1195      | 1             | 355          |
|              | 1196      | 1             | 1            |
|              | 1200      | 1             | 33           |

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

| Config ID | ORF ID | Start (nt) | Stop (nt) |
|-----------|--------|------------|-----------|
| 1203      | 2      | 129        | 464       |
| 1222      | 2      | 105        | 401       |
| 1232      | 1      | 1          | 387       |
| 1240      | 1      | 2          | 175       |
| 1247      | 1      | 520        | 311       |
| 1271      | 1      | 412        | 221       |
| 1286      | 1      | 2          | 595       |
| 1295      | 1      | 1          | 165       |
| 1306      | 1      | 367        | 185       |
| 1314      | 2      | 158        | 631       |
| 1316      | 1      | 58         | 570       |
| 1359      | 1      | 384        | 193       |
| 1370      | 1      | 1          | 402       |
| 1371      | 1      | 1          | 345       |
| 1374      | 1      | 710        | 357       |
| 1378      | 1      | 2          | 400       |
| 1392      | 1      | 3          | 413       |
| 1411      | 1      | 202        | 432       |
| 1433      | 1      | 331        | 167       |
| 1450      | 1      | 2          | 256       |
| 1453      | 1      | 295        | 149       |
| 1471      | 1      | 721        | 398       |
| 1477      | 1      | 869        | 639       |
| 1502      | 1      | 794        | 399       |
| 1518      | 1      | 126        | 449       |
| 1534      | 1      | 283        | 143       |
| 1546      | 1      | 3          | 401       |

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|-----------|--------|------------|-----------|
| 1547      | 1      | 506        | 255       |
| 1583      | 1      | 3          | 350       |
| 1587      | 1      | 3          | 563       |
| 1602      | 2      | 170        | 679       |
| 1629      | 1      | 1          | 402       |
| 1665      | 1      | 468        | 235       |
| 1760      | 1      | 625        | 314       |
| 1762      | 1      | 3          | 200       |
| 1876      | 2      | 119        | 286       |
| 1895      | 1      | 2          | 379       |
| 1931      | 1      | 798        | 400       |
| 1976      | 2      | 715        | 383       |
| 2055      | 2      | 252        | 401       |
| 2056      | 1      | 331        | 167       |
| 2150      | 1      | 523        | 263       |
| 2157      | 1      | 794        | 399       |
| 2164      | 1      | 564        | 283       |
| 2175      | 1      | 218        | 400       |
| 2212      | 1      | 492        | 331       |
| 2338      | 1      | 732        | 367       |
| 2342      | 1      | 3          | 167       |
| 2352      | 1      | 330        | 166       |
| 2352      | 2      | 622        | 398       |
| 2355      | 1      | 47         | 352       |
| 2356      | 1      | 679        | 341       |
| 2359      | 1      | 301        | 152       |
| 2421      | 1      | 296        | 150       |

5. aureus - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|-----------|--------|------------|-----------|
| 3046      | 1      | 367        | 185       |
| 3049      | 1      | 553        | 278       |
| 3050      | 1      | 3          | 314       |
| 3052      | 1      | 504        | 251       |
| 3065      | 1      | 2          | 157       |
| 3070      | 1      | 357        | 190       |
| 3075      | 1      | 440        | 222       |
| 3080      | 1      | 1          | 245       |
| 3092      | 1      | 320        | 162       |
| 3093      | 1      | 411        | 230       |
| 3100      | 1      | 52         | 237       |
| 3103      | 1      | 47         | 288       |
| 3118      | 1      | 344        | 174       |
| 3123      | 1      | 2          | 145       |
| 3127      | 1      | 1          | 147       |
| 3138      | 1      | 336        | 169       |
| 3142      | 1      | 388        | 203       |
| 3144      | 1      | 664        | 386       |
| 3151      | 1      | 337        | 170       |
| 3155      | 2      | 202        | 384       |
| 3168      | 1      | 12         | 176       |
| 3205      | 1      | 288        | 145       |
| 3282      | 1      | 1          | 150       |
| 3303      | 2      | 239        | 400       |
| 3371      | 2      | 211        | 399       |
| 3536      | 1      | 2          | 148       |
| 3538      | 2      | 36         | 401       |

TABLE 3

S. aureus - putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|-----------|--------|------------|-----------|
| 3046      | 1      | 367        | 185       |
| 3049      | 1      | 553        | 278       |
| 3050      | 1      | 3          | 314       |
| 3052      | 1      | 504        | 253       |
| 3065      | 1      | 2          | 157       |
| 3070      | 1      | 357        | 190       |
| 3075      | 1      | 440        | 222       |
| 3080      | 1      | 1          | 285       |
| 3082      | 1      | 320        | 162       |
| 3093      | 1      | 411        | 250       |
| 3100      | 1      | 52         | 237       |
| 3103      | 1      | 47         | 288       |
| 3118      | 1      | 344        | 174       |
| 3123      | 1      | 2          | 145       |
| 3127      | 1      | 1          | 147       |
| 3136      | 1      | 336        | 169       |
| 3142      | 1      | 388        | 203       |
| 3144      | 1      | 664        | 386       |
| 3151      | 1      | 337        | 170       |
| 3155      | 2      | 202        | 384       |
| 3168      | 1      | 12         | 176       |
| 3205      | 1      | 288        | 145       |
| 3282      | 1      | 1          | 150       |
| 3303      | 2      | 239        | 400       |
| 3371      | 2      | 211        | 399       |
| 3558      | 1      | 2          | 148       |
| 3558      | 2      | 36         | 401       |

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|-----------|--------|------------|-----------|
| 3568      | 1      | 751        | 377       |
| 3595      | 1      | 757        | 380       |
| 3618      | 1      | 2          | 238       |
| 3618      | 2      | 130        | 402       |
| 3622      | 1      | 86         | 358       |
| 3622      | 2      | 664        | 398       |
| 3642      | 1      | 876        | 439       |
| 3649      | 1      | 781        | 398       |
| 3651      | 1      | 625        | 314       |
| 3664      | 1      | 467        | 637       |
| 3674      | 1      | 55         | 402       |
| 3677      | 1      | 619        | 311       |
| 3704      | 1      | 1          | 402       |
| 3726      | 1      | 535        | 269       |
| 3765      | 1      | 510        | 256       |
| 3779      | 1      | 554        | 357       |
| 3794      | 1      | 266        | 135       |
| 3794      | 2      | 667        | 377       |
| 3796      | 2      | 638        | 375       |
| 3801      | 1      | 474        | 262       |
| 3806      | 1      | 453        | 298       |
| 3807      | 1      | 42         | 389       |
| 3815      | 1      | 798        | 400       |
| 3827      | 1      | 3          | 320       |
| 3842      | 1      | 781        | 392       |
| 3853      | 1      | 671        | 399       |
| 3855      | 1      | 1          | 324       |

TABLE 3

S. aureus - putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|-----------|--------|------------|-----------|
| 3857      | 1      | 2          | 235       |
| 3861      | 1      | 590        | 297       |
| 3865      | 1      | 695        | 399       |
| 3897      | 1      | 3          | 173       |
| 3897      | 2      | 143        | 400       |
| 3898      | 2      | 225        | 401       |
| 3921      | 2      | 103        | 342       |
| 3927      | 1      | 70         | 375       |
| 3930      | 1      | 76         | 234       |
| 3946      | 2      | 651        | 382       |
| 3951      | 2      | 105        | 377       |
| 3965      | 1      | 646        | 344       |
| 3973      | 1      | 795        | 400       |
| 3981      | 1      | 3          | 311       |
| 3998      | 1      | 3          | 356       |
| 4001      | 1      | 481        | 296       |
| 4003      | 1      | 90         | 335       |
| 4018      | 1      | 2          | 259       |
| 4018      | 2      | 186        | 401       |
| 4021      | 1      | 1          | 345       |
| 4043      | 1      | 3          | 344       |
| 4054      | 1      | 3          | 344       |
| 4066      | 1      | 1          | 150       |
| 4070      | 1      | 1          | 324       |
| 4072      | 2      | 187        | 390       |
| 4073      | 1      | 1          | 285       |
| 4077      | 2      | 127        | 372       |

TABLE 3

*S. aureus* - Putative coding regions of novel proteins not similar to known proteins

| Contig | ORF | Start | Stop |
|--------|-----|-------|------|
| Id     | Int | Int   | Int  |
| 4083   | 1   | 3     | 359  |
| 4090   | 1   | 27    | 368  |
| 4101   | 1   | 103   | 297  |
| 4105   | 1   | 1     | 306  |
| 4107   | 1   | 570   | 286  |
| 4119   | 1   | 629   | 339  |
| 4121   | 1   | 760   | 372  |
| 4123   | 1   | 3     | 230  |
| 4127   | 1   | 3     | 341  |
| 4128   | 1   | 2     | 331  |
| 4130   | 1   | 768   | 415  |
| 4146   | 1   | 97    | 381  |
| 4157   | 1   | 3     | 286  |
| 4186   | 1   | 505   | 254  |
| 4224   | 1   | 510   | 256  |
| 4239   | 1   | 1     | 348  |
| 4242   | 1   | 709   | 356  |
| 4252   | 1   | 589   | 296  |
| 4253   | 1   | 1     | 174  |
| 4256   | 1   | 568   | 323  |
| 4258   | 2   | 498   | 334  |
| 4267   | 1   | 284   | 144  |
| 4271   | 1   | 2     | 304  |
| 4287   | 1   | 303   | 163  |
| 4289   | 1   | 471   | 319  |
| 4302   | 1   | 153   | 305  |
| 4304   | 1   | 1     | 186  |

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|-----------|--------|------------|-----------|
| 4304      | 2      | 96         | 314       |
| 4306      | 1      | 2          | 151       |
| 4318      | 1      | 576        | 289       |
| 4322      | 1      | 5          | 148       |
| 4331      | 1      | 439        | 221       |
| 4331      | 2      | 528        | 364       |
| 4338      | 1      | 728        | 399       |
| 4346      | 1      | 471        | 277       |
| 4357      | 2      | 117        | 311       |
| 4373      | 1      | 2          | 268       |
| 4381      | 1      | 574        | 326       |
| 4384      | 1      | 614        | 309       |
| 4397      | 1      | 9          | 311       |
| 4402      | 1      | 1          | 249       |
| 4403      | 1      | 606        | 328       |
| 4406      | 1      | 3          | 317       |
| 4411      | 1      | 2          | 280       |
| 4411      | 2      | 697        | 398       |
| 4412      | 1      | 2          | 364       |
| 4418      | 1      | 3          | 230       |
| 4424      | 1      | 601        | 398       |
| 4443      | 1      | 427        | 215       |
| 4471      | 1      | 643        | 323       |
| 4478      | 1      | 540        | 271       |
| 4482      | 1      | 50         | 289       |
| 4489      | 1      | 601        | 302       |
| 4491      | 1      | 12         | 206       |

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

| Contig<br>ID | ORF<br>ID | Start<br>(nt) | Stop<br>(nt) |
|--------------|-----------|---------------|--------------|
| 4495         | 1         | 3             | 179          |
| 4496         | 1         | 300           | 252          |
| 4500         | 1         | 130           | 306          |
| 4511         | 1         | 493           | 248          |
| 4518         | 1         | 1             | 246          |
| 4526         | 1         | 480           | 241          |
| 4527         | 1         | 2             | 163          |
| 4532         | 1         | 3             | 239          |
| 4542         | 1         | 11            | 175          |
| 4567         | 1         | 36            | 200          |
| 4573         | 1         | 1             | 231          |
| 4578         | 1         | 642           | 322          |
| 4619         | 1         | 1             | 180          |
| 4620         | 1         | 349           | 176          |
| 4662         | 1         | 1             | 246          |
| 4669         | 1         | 2             | 157          |
| 4680         | 1         | 28            | 183          |
| 4690         | 1         | 344           | 174          |

Table 4

|    | ORF    | SEQ ID NO | BLAST                         | Antigenic Regions |          |          |          |
|----|--------|-----------|-------------------------------|-------------------|----------|----------|----------|
|    |        |           | HOMOLOG                       | Region 1          | Region 2 | Region 3 | Region 4 |
| 5  | 168_6  | 5192      | lipoprotein                   | 36-45             | 84-103   | 152-161  | 176-185  |
|    | 238_1  | 5193      | chrA                          | 21-39             | 48-58    | 84-95    | 232-249  |
|    | 51_2   | 5194      | OppB gene product (B. sub     | 20-36             | 70-79    | 100-112  | 121-131  |
|    | 278_3  | 5195      | lipoprotein 1                 | 20-29             | 59-73    | 85-97    | 162-171  |
| 10 | 276_2  | 5196      | lipoprotein                   | 21-33             | 65-74    | 177-186  | 211-220  |
|    | 45_4   | 5197      | ProX                          | 28-37             | 59-69    | 85-100   | 120-129  |
|    | 316_8  | 5198      | hypothetical protein          | 45-54             | 88-97    | 182-192  | 243-253  |
|    | 154_15 | 5199      | unknown                       | 31-40             | 48-58    | 79-88    | 95-104   |
|    | 228_3  | 5200      | unknown                       | 25-38             | 40-52    | 64-74    | 80-89    |
| 15 | 228_6  | 5201      | unknown                       | 29-41             | 89-101   | 128-143  | 173-184  |
|    | 50_1   | 5202      | unknown                       | 21-33             | 52-61    | 168-182  | 197-206  |
|    | 112_7  | 5203      | iron-binding periplasmic      | 21-31             | 58-67    | 92-101   | 111-120  |
|    | 442_1  | 5204      | unknown                       | 30-39             | 91-100   | 122-137  | 182-192  |
|    | 66_2   | 5205      | unknown                       | 50-59             | 104-116  | 127-136  | 167-182  |
| 20 | 304_2  | 5206      | Q-binding periplasmic         | 19-28             | 48-57    | 75-84    | 103-116  |
|    | 44_1   | 5207      | hypothetical protein          | 27-36             | 86-95    | 129-138  | 192-201  |
|    | 161_4  | 5208      | SphX                          | 27-44             | 149-161  | 166-175  | 201-210  |
|    | 46_5   | 5209      | cmpC (permease)               | 21-33             | 61-70    | 83-92    | 100-109  |
|    | 942_1  | 5210      | traH [Plasmid pSK41]          | 83-92             | 109-118  | 127-142  |          |
| 25 | 5_4    | 5211      | ORF (S. aureus)               | 12-22             | 87-96    | 111-120  | 151-160  |
|    | 20_4   | 5212      | peptidoglycan hydrolase (S.   | 24-34             | 129-138  | 141-150  | 161-171  |
|    | 328_2  | 5213      | lipoprotein (H. flu)          | 81-90             | 123-133  | 290-299  |          |
|    | 520_2  | 5214      | fibronectin binding protein   | 44-54             | 63-79    | 81-90    | 95-110   |
|    | 771_1  | 5215      | emm1 gene product (S. pyo     | 30-39             | 65-82    | 96-106   | 112-121  |
|    | 999_1  | 5216      | predicted trithorax prot. (D  | 7-16              | 120-129  | 157-166  |          |
| 30 | 853_1  | 5217      | ORF2136 (Marchantia polyr     | 43-52             | 88-97    | 102-111  |          |
|    | 287_1  | 5218      | psaA homolog                  | 13-22             | 28-44    | 72-82    | 114-124  |
|    | 288_2  | 5219      | cell wall enzyme              | 14-23             | 89-98    |          |          |
|    | 596_2  | 5220      | penicillin binding protein 2b | 40-49             | 59-68    | 76-87    | 106-115  |
|    | 217_5  | 5221      | fibronectin/fibrinogen bindi  | 28-37             | 40-49    | 62-71    | 93-111   |
| 35 | 217_6  | 5222      | fibronectin/fibrinogen bp     | 10-19             | 31-40    | 54-62    | 73-92    |
|    | 528_3  | 5223      | myosin cross reactive prote   | 4-13              | 29-47    | 60-73    | 90-99    |
|    | 171_11 | 5224      | EF                            | 20-31             | 91-110   |          |          |
|    | 63_4   | 5225      | penicillin binding protein 2b | 12-21             | 59-68    | 95-104   |          |
|    | 353_2  | 5226      |                               | 46-55             | 62-71    |          |          |
| 40 | 743_1  | 5227      | 29 kDa protein in fimA regi   | 23-32             | 68-79    | 94-103   | 175-184  |
|    | 342_4  | 5228      | Twitching motility            | 10-19             | 48-60    | 83-92    | 111-121  |
|    | 69_3   | 5229      | arabinogalactan protein       | 97-106            | 132-141  | 158-167  | 180-189  |
|    | 70_6   | 5230      | nodulin                       | 36-45             | 48-57    | 137-160  | 179-188  |
|    | 129_2  | 5231      | glycerol diester phosphodie   | 8-17              | 41-50    | 55-74    | 97-106   |
|    | 58_5   | 5232      | PBP (S. aureus)               | 26-35             | 70-79    | 117-126  | 152-161  |
| 45 | 188_3  | 5233      | MHC class II analog (S. aure  | 72-81             | 94-103   | 115-124  | 136-145  |
|    | 236_6  | 5234      | histidine kinase domain (Dic  | 24-33             | 52-67    | 81-94    | 106-121  |
|    | 310_8  | 5235      | clumping factor (S. aureus)   | 59-71             | 77-86    | 93-102   | 118-127  |
|    | 601_1  | 5236      | novel antigen/ORF2 (S. au     | 45-54             | 91-104   | 108-117  | 186-195  |
|    | 544_3  | 5237      | ORF YJR151c (S. cerevisae     | 76-90             | 101-111  | 131-140  | 154-164  |
| 50 | 662_1  | 5238      | MHC class II analog (S. aure  | 22-32             | 71-80    | 89-98    | 114-122  |
|    | 87_7   | 5239      | 5' nucleotidase precursor ('  | 29-45             | 62-71    | 105-114  | 125-137  |
|    | 120_1  | 5240      | B65G gene product (B. sub     | 102-111           |          |          |          |

Table 4

| ORF    | Antigenic Regions |          | (cont)   |          |          |           |
|--------|-------------------|----------|----------|----------|----------|-----------|
|        | Region 5          | Region 6 | Region 7 | Region 8 | Region 9 | Region 10 |
| 168_6  | 244-272           | 303-315  |          |          |          |           |
| 238_1  | 260-269           | 291-301  | 308-317  |          |          |           |
| 51_2   | 140-152           | 188-208  | 211-220  | 256-266  | 273-283  |           |
| 278_3  | 198-209           |          |          |          |          |           |
| 276_2  | 255-268           |          |          |          |          |           |
| 45_4   | 177-199           | 221-230  | 234-243  | 268-279  | 284-293  | 304-313   |
| 316_8  |                   |          |          |          |          |           |
| 154_15 | 148-157           | 177-187  | 202-211  |          |          |           |
| 228_3  | 101-119           | 139-154  | 166-181  |          |          |           |
| 228_6  |                   |          |          |          |          |           |
| 50_1   |                   |          |          |          |          |           |
| 112_7  | 136-149           | 197-211  | 218-229  | 253-273  |          |           |
| 442_1  | 199-210           | 247-257  | 264-277  | 287-309  |          |           |
| 66_2   |                   |          |          |          |          |           |
| 304_2  | 178-187           | 250-259  |          |          |          |           |
| 44_1   |                   |          |          |          |          |           |
| 161_4  |                   |          |          |          |          |           |
| 46_5   | 131-141           | 162-176  | 206-215  | 243-252  | 264-273  | 285-294   |
| 942_1  |                   |          |          |          |          |           |
| 5_4    | 189-205           | 230-239  | 246-264  | 301-318  | 340-354  | 378-387   |
| 20_4   | 202-212           | 217-234  | 260-275  | 314-336  | 366-373  | 380-391   |
| 328_2  |                   |          |          |          |          |           |
| 520_2  |                   |          |          |          |          |           |
| 771_1  | 145-154           |          |          |          |          |           |
| 999_1  |                   |          |          |          |          |           |
| 853_1  |                   |          |          |          |          |           |
| 287_1  | 154-164           |          |          |          |          |           |
| 288_2  |                   |          |          |          |          |           |
| 596_2  | 121-130           |          |          |          |          |           |
| 217_5  | 244-253           | 259-268  | 288-297  | 302-311  |          |           |
| 217_6  | 144-158           | 174-183  | 188-197  | 207-216  | 226-242  |           |
| 528_3  |                   |          |          |          |          |           |
| 171_11 |                   |          |          |          |          |           |
| 63_4   |                   |          |          |          |          |           |
| 353_2  |                   |          |          |          |          |           |
| 743_1  | 197-207           |          |          |          |          |           |
| 342_4  |                   |          |          |          |          |           |
| 69_3   | 195-211           |          |          |          |          |           |
| 70_6   | 206-215           | 263-272  | 291-301  | 331-340  | 358-371  | 390-414   |
| 129_2  | 117-127           | 141-157  | 168-183  | 202-211  | 222-231  | 261-270   |
| 58_5   | 184-203           | 260-269  | 275-299  | 330-344  | 372-381  | 424-433   |
| 188_3  |                   |          |          |          |          |           |
| 236_6  | 138-147           | 163-172  | 187-198  | 244-261  | 268-278  | 308-317   |
| 310_8  | 131-140           | 144-153  | 177-186  | 190-199  | 204-213  | 216-227   |
| 601_1  | 208-218           |          |          |          |          |           |
| 544_3  | 170-179           | 184-193  | 224-235  | 274-287  | 327-336  | 352-361   |
| 662_1  |                   |          |          |          |          |           |
| 87_7   |                   |          |          |          |          |           |
| 120_1  |                   |          |          |          |          |           |

Table 4

|    | ORF    | Antigenic Regions |           | (cont)    |           |           |           |
|----|--------|-------------------|-----------|-----------|-----------|-----------|-----------|
|    |        | Region 11         | Region 12 | Region 13 | Region 14 | Region 15 | Region 16 |
| 5  | 168_6  |                   |           |           |           |           |           |
|    | 238_1  |                   |           |           |           |           |           |
|    | 51_2   |                   |           |           |           |           |           |
| 10 | 278_3  |                   |           |           |           |           |           |
|    | 276_2  |                   |           |           |           |           |           |
|    | 45_4   |                   |           |           |           |           |           |
|    | 316_8  |                   |           |           |           |           |           |
|    | 154_15 |                   |           |           |           |           |           |
| 15 | 228_3  |                   |           |           |           |           |           |
|    | 228_6  |                   |           |           |           |           |           |
|    | 50_1   |                   |           |           |           |           |           |
|    | 112_7  |                   |           |           |           |           |           |
|    | 442_1  |                   |           |           |           |           |           |
| 20 | 66_2   |                   |           |           |           |           |           |
|    | 304_2  |                   |           |           |           |           |           |
|    | 44_1   |                   |           |           |           |           |           |
|    | 161_4  |                   |           |           |           |           |           |
|    | 46_5   | 306-315           |           |           |           |           |           |
| 25 | 942_1  |                   |           |           |           |           |           |
|    | 5_4    | 393-407           | 416-426   | 456-465   |           |           |           |
|    | 20_4   | 396-405           | 410-419   | 461-481   |           |           |           |
|    | 328_2  |                   |           |           |           |           |           |
|    | 520_2  |                   |           |           |           |           |           |
| 30 | 771_1  |                   |           |           |           |           |           |
|    | 999_1  |                   |           |           |           |           |           |
|    | 853_1  |                   |           |           |           |           |           |
|    | 287_1  |                   |           |           |           |           |           |
|    | 288_2  |                   |           |           |           |           |           |
| 35 | 596_2  |                   |           |           |           |           |           |
|    | 217_5  |                   |           |           |           |           |           |
|    | 217_6  |                   |           |           |           |           |           |
|    | 528_3  |                   |           |           |           |           |           |
|    | 171_11 |                   |           |           |           |           |           |
| 40 | 63_4   |                   |           |           |           |           |           |
|    | 353_2  |                   |           |           |           |           |           |
|    | 743_1  |                   |           |           |           |           |           |
|    | 342_4  |                   |           |           |           |           |           |
|    | 69_3   |                   |           |           |           |           |           |
| 45 | 70_6   | 453-471           | 506-515   |           |           |           |           |
|    | 129_2  | 296-315           |           |           |           |           |           |
|    | 58_5   |                   |           |           |           |           |           |
|    | 188_3  |                   |           |           |           |           |           |
|    | 236_6  | 358-377           | 410-423   | 428-439   | 442-457   | 467-476   | 480-493   |
| 50 | 310_8  | 238-251           | 256-275   | 281-290   | 296-310   | 314-333   | 338-347   |
|    | 601_1  |                   |           |           |           |           |           |
|    | 544_3  |                   |           |           |           |           |           |
|    | 662_1  |                   |           |           |           |           |           |
| 55 | 87_7   |                   |           |           |           |           |           |
|    | 120_1  |                   |           |           |           |           |           |

## EP 0 786 519 A2

Table 4

|    | ORF    | Antigenic Regions (cont) |           |           |           |           |
|----|--------|--------------------------|-----------|-----------|-----------|-----------|
|    |        | Region 17                | Region 18 | Region 19 | Region 20 | Region 21 |
| 5  | 168_6  |                          |           |           |           |           |
|    | 238_1  |                          |           |           |           |           |
|    | 51_2   |                          |           |           |           |           |
| 10 | 278_3  |                          |           |           |           |           |
|    | 276_2  |                          |           |           |           |           |
|    | 45_4   |                          |           |           |           |           |
|    | 316_8  |                          |           |           |           |           |
|    | 154_15 |                          |           |           |           |           |
| 15 | 228_3  |                          |           |           |           |           |
|    | 228_6  |                          |           |           |           |           |
|    | 50_1   |                          |           |           |           |           |
|    | 112_7  |                          |           |           |           |           |
|    | 442_1  |                          |           |           |           |           |
| 20 | 66_2   |                          |           |           |           |           |
|    | 304_2  |                          |           |           |           |           |
|    | 44_1   |                          |           |           |           |           |
|    | 161_4  |                          |           |           |           |           |
|    | 46_5   |                          |           |           |           |           |
| 25 | 942_1  |                          |           |           |           |           |
|    | 5_4    |                          |           |           |           |           |
|    | 20_4   |                          |           |           |           |           |
|    | 328_2  |                          |           |           |           |           |
|    | 520_2  |                          |           |           |           |           |
| 30 | 771_1  |                          |           |           |           |           |
|    | 999_1  |                          |           |           |           |           |
|    | 853_1  |                          |           |           |           |           |
|    | 287_1  |                          |           |           |           |           |
|    | 288_2  |                          |           |           |           |           |
| 35 | 596_2  |                          |           |           |           |           |
|    | 217_5  |                          |           |           |           |           |
|    | 217_6  |                          |           |           |           |           |
|    | 528_3  |                          |           |           |           |           |
|    | 171_11 |                          |           |           |           |           |
| 40 | 63_4   |                          |           |           |           |           |
|    | 353_2  |                          |           |           |           |           |
|    | 743_1  |                          |           |           |           |           |
|    | 342_4  |                          |           |           |           |           |
|    | 69_3   |                          |           |           |           |           |
| 45 | 70_6   |                          |           |           |           |           |
|    | 129_2  |                          |           |           |           |           |
|    | 58_5   |                          |           |           |           |           |
|    | 188_3  |                          |           |           |           |           |
|    | 236_6  |                          |           |           |           |           |
| 50 | 310_8  | 357-366                  | 370-379   | 429-438   | 443-452   | 478-487   |
|    | 601_1  |                          |           |           |           | 551-560   |
|    | 544_3  |                          |           |           |           |           |
|    | 662_1  |                          |           |           |           |           |
| 55 | 87_7   |                          |           |           |           |           |
|    | 120_1  |                          |           |           |           |           |

Table 4

|    | ORF    | Antigenic Regions (cont) |           |           |           |           |           |
|----|--------|--------------------------|-----------|-----------|-----------|-----------|-----------|
|    |        | Region 23                | Region 24 | Region 25 | Region 26 | Region 27 | Region 28 |
|    | 168_6  |                          |           |           |           |           |           |
|    | 238_1  |                          |           |           |           |           |           |
|    | 51_2   |                          |           |           |           |           |           |
| 10 | 278_3  |                          |           |           |           |           |           |
|    | 276_2  |                          |           |           |           |           |           |
|    | 45_4   |                          |           |           |           |           |           |
|    | 316_8  |                          |           |           |           |           |           |
|    | 154_15 |                          |           |           |           |           |           |
| 15 | 228_3  |                          |           |           |           |           |           |
|    | 228_6  |                          |           |           |           |           |           |
|    | 50_1   |                          |           |           |           |           |           |
|    | 112_7  |                          |           |           |           |           |           |
|    | 442_1  |                          |           |           |           |           |           |
| 20 | 66_2   |                          |           |           |           |           |           |
|    | 304_2  |                          |           |           |           |           |           |
|    | 44_1   |                          |           |           |           |           |           |
|    | 161_4  |                          |           |           |           |           |           |
|    | 46_5   |                          |           |           |           |           |           |
| 25 | 942_1  |                          |           |           |           |           |           |
|    | 5_1    |                          |           |           |           |           |           |
|    | 20_4   |                          |           |           |           |           |           |
|    | 328_2  |                          |           |           |           |           |           |
|    | 520_2  |                          |           |           |           |           |           |
| 30 | 771_1  |                          |           |           |           |           |           |
|    | 999_1  |                          |           |           |           |           |           |
|    | 853_1  |                          |           |           |           |           |           |
|    | 287_1  |                          |           |           |           |           |           |
|    | 288_2  |                          |           |           |           |           |           |
| 35 | 596_2  |                          |           |           |           |           |           |
|    | 217_5  |                          |           |           |           |           |           |
|    | 217_6  |                          |           |           |           |           |           |
|    | 528_3  |                          |           |           |           |           |           |
|    | 171_11 |                          |           |           |           |           |           |
| 40 | 63_4   |                          |           |           |           |           |           |
|    | 353_2  |                          |           |           |           |           |           |
|    | 743_1  |                          |           |           |           |           |           |
|    | 342_4  |                          |           |           |           |           |           |
|    | 69_3   |                          |           |           |           |           |           |
| 45 | 70_6   |                          |           |           |           |           |           |
|    | 129_2  |                          |           |           |           |           |           |
|    | 58_5   |                          |           |           |           |           |           |
|    | 188_3  |                          |           |           |           |           |           |
|    | 236_6  |                          |           |           |           |           |           |
| 50 | 310_8  | 622-632                  | 670-685   | 708-718   | 823-836   | 858-867   | 877-886   |
|    | 601_1  |                          |           |           |           |           |           |
|    | 544_3  |                          |           |           |           |           |           |
|    | 662_1  |                          |           |           |           |           |           |
|    | 87_7   |                          |           |           |           |           |           |
| 55 | 120_1  |                          |           |           |           |           |           |

Table 4

|    | ORF    | Antigenic Regions |           | (cont) |
|----|--------|-------------------|-----------|--------|
|    |        | Region 29         | Region 30 |        |
| 5  | 168_6  |                   |           |        |
|    | 238_1  |                   |           |        |
|    | 51_2   |                   |           |        |
| 10 | 278_3  |                   |           |        |
|    | 276_2  |                   |           |        |
|    | 45_4   |                   |           |        |
|    | 316_8  |                   |           |        |
|    | 154_15 |                   |           |        |
| 15 | 228_3  |                   |           |        |
|    | 228_6  |                   |           |        |
|    | 50_1   |                   |           |        |
|    | 112_7  |                   |           |        |
|    | 442_1  |                   |           |        |
| 20 | 66_2   |                   |           |        |
|    | 304_2  |                   |           |        |
|    | 44_1   |                   |           |        |
|    | 161_4  |                   |           |        |
|    | 46_5   |                   |           |        |
| 25 | 942_1  |                   |           |        |
|    | 5_4    |                   |           |        |
|    | 20_4   |                   |           |        |
|    | 328_2  |                   |           |        |
|    | 520_2  |                   |           |        |
| 30 | 771_1  |                   |           |        |
|    | 999_1  |                   |           |        |
|    | 853_1  |                   |           |        |
|    | 287_1  |                   |           |        |
|    | 288_2  |                   |           |        |
| 35 | 596_2  |                   |           |        |
|    | 217_5  |                   |           |        |
|    | 217_6  |                   |           |        |
|    | 528_3  |                   |           |        |
|    | 171_11 |                   |           |        |
| 40 | 63_4   |                   |           |        |
|    | 353_2  |                   |           |        |
|    | 743_1  |                   |           |        |
|    | 342_4  |                   |           |        |
|    | 69_3   |                   |           |        |
| 45 | 70_6   |                   |           |        |
|    | 129_2  |                   |           |        |
|    | 58_5   |                   |           |        |
|    | 188_3  |                   |           |        |
|    | 236_6  |                   |           |        |
| 50 | 310_8  |                   |           |        |
|    | 601_1  |                   |           |        |
|    | 544_3  |                   |           |        |
|    | 662_1  |                   |           |        |
| 55 | 87_7   |                   |           |        |
|    | 120_1  |                   |           |        |

Table 4

| ORF    | BLAST   |                               | Antigenic Regions |          |          |          |
|--------|---------|-------------------------------|-------------------|----------|----------|----------|
|        | HOMOLOG |                               | Region 1          | Region 2 | Region 3 | Region 4 |
| 46_1   | 5241    | aldehyde dehydrogenase        | 8-17              | 36-52    | 83-96    | 112-121  |
| 63_4   | 5242    | glycerol ester hydrolase (P.  | 9-26              | 57-73    | 93-107   | 123-133  |
| 174_6  | 5243    | ketopantoate hydroxymeth      | 71-80             | 203-212  | 242-254  | 265-274  |
| 206_16 | 5244    | ornithine acetyltransferase   | 1-10              | 34-43    | 54-63    | 194-210  |
| 267_1  | 5245    | NaH-antiporter protein (E. t  | 120-129           | 332-347  | 398-408  |          |
| 322_1  | 5246    | acriflavin resistance protein | 58-75             | 153-164  | 203-231  | 264-284  |
| 415_2  | 5247    | transport ATP-binding prote   | 108-126           | 218-227  | 298-308  | 315-334  |
| 214_3  | 5248    | 2-nitropropane dioxygenase    | 123-136           | 216-233  | 283-292  | 297-306  |
| 587_3  | 5249    | clumping factor               | 5-14              | 43-54    | 59-68    | 76-95    |
| 685_1  | 5250    | signal peptidase              | 59-68             | 72-81    | 86-95    | 99-108   |
| 54_3   | 5251    | fibronectin binding protein I | 23-32             | 37-46    | 50-59    | 89-98    |
| 54_4   | 5252    | fibronectin binding protein I | 43-52             | 66-75    | 95-104   | 147-156  |
| 54_5   | 5253    | fibronectin binding protein I | 49-60             | 81-90    |          |          |
| 54_6   | 5254    | fibronectin binding protein I | 55-71             | 82-97    | 139-158  | 175-186  |
| 328_1  | 5255    | lipoprotein (H. flu)          | 11-20             | 61-70    | 96-105   |          |

Table 4

| ORF    | Antigenic Regions (cont) |          |          |          |          |           |
|--------|--------------------------|----------|----------|----------|----------|-----------|
|        | Region 5                 | Region 6 | Region 7 | Region 8 | Region 9 | Region 10 |
| 46_1   | 215-242                  | 333-352  | 376-385  | 416-432  | 471-487  |           |
| 63_4   | 145-154                  | 191-202  | 212-223  | 245-265  | 274-283  | 291-300   |
| 174_6  |                          |          |          |          |          |           |
| 206_16 | 239-259                  | 275-284  |          |          |          |           |
| 267_1  |                          |          |          |          |          |           |
| 322_1  | 298-319                  | 350-359  |          |          |          |           |
| 415_2  | 344-353                  | 371-380  | 395-404  | 456-465  | 486-495  | 518-527   |
| 214_3  | 318-337                  | 365-375  |          |          |          |           |
| 587_3  | 106-115                  | 142-151  | 156-166  | 173-182  | 186-198  | 204-213   |
| 685_1  | 113-122                  | 130-145  |          |          |          |           |
| 54_3   | 128-138                  | 185-194  | 217-226  | 251-260  | 268-277  | 295-305   |
| 54_4   | 175-188                  | 191-200  | 203-212  | 220-229  |          |           |
| 54_5   |                          |          |          |          |          |           |
| 54_6   | 220-230                  | 287-304  | 317-326  | 344-353  | 364-373  | 378-387   |
| 328_1  |                          |          |          |          |          |           |

Table 4

| ORF    | Antigenic Regions |           |           | (cont)    |           |           |
|--------|-------------------|-----------|-----------|-----------|-----------|-----------|
|        | Region 11         | Region 12 | Region 13 | Region 14 | Region 15 | Region 17 |
| 46_1   |                   |           |           |           |           |           |
| 63_4   | 306-315           | 319-328   | 366-376   | 395-420   | 453-462   | 467-476   |
| 174_6  |                   |           |           |           |           |           |
| 206_16 |                   |           |           |           |           |           |
| 267_1  |                   |           |           |           |           |           |
| 322_1  |                   |           |           |           |           |           |
| 415_2  | 539-555           |           |           |           |           |           |
| 214_3  |                   |           |           |           |           |           |
| 587_3  | 217-226           | 278-287   | 318-327   | 332-342   | 351-360   | 377-386   |
| 685_1  |                   |           |           |           |           |           |
| 54_3   | 316-325           | 329-345   | 355-372   | 387-396   | 416-425   | 438-448   |
| 54_4   |                   |           |           |           |           |           |
| 54_5   |                   |           |           |           |           |           |
| 54_6   | 396-407           | 427-436   | 514-531   | 541-550   | 569-578   | 612-622   |
| 328_1  |                   |           |           |           |           |           |

Table 4

| ORF    | Antigenic Regions |           |           | (cont)    |           |           |
|--------|-------------------|-----------|-----------|-----------|-----------|-----------|
|        | Region 18         | Region 19 | Region 20 | Region 21 | Region 22 | Region 23 |
| 46_1   |                   |           |           |           |           |           |
| 63_4   | 485-500           | 513-525   |           |           |           |           |
| 174_6  |                   |           |           |           |           |           |
| 206_16 |                   |           |           |           |           |           |
| 267_1  |                   |           |           |           |           |           |
| 322_1  |                   |           |           |           |           |           |
| 415_2  |                   |           |           |           |           |           |
| 214_3  |                   |           |           |           |           |           |
| 587_3  | 396-405           | 426-442   | 459-470   | 485-494   | 505-514   | 531-562   |
| 685_1  |                   |           |           |           |           |           |
| 54_3   | 455-462           | 472-491   | 517-536   |           |           |           |
| 54_4   |                   |           |           |           |           |           |
| 54_5   |                   |           |           |           |           |           |
| 54_6   | 639-648           | 673-681   | 703-715   | 723-732   | 749-760   | 772-788   |
| 328_1  |                   |           |           |           |           |           |

Table 4

| ORF    | Antigenic Regions (cont) |           |           |           |           |           |
|--------|--------------------------|-----------|-----------|-----------|-----------|-----------|
|        | Region 24                | Region 25 | Region 26 | Region 27 | Region 28 | Region 29 |
| 46_1   |                          |           |           |           |           |           |
| 63_4   |                          |           |           |           |           |           |
| 174_6  |                          |           |           |           |           |           |
| 206_16 |                          |           |           |           |           |           |
| 267_1  |                          |           |           |           |           |           |
| 322_1  |                          |           |           |           |           |           |
| 415_2  |                          |           |           |           |           |           |
| 214_3  |                          |           |           |           |           |           |
| 587_3  | 567-578                  | 584-601   | 607-840   | 844-854   | 858-870   | 877-886   |
| 685_1  |                          |           |           |           |           |           |
| 54_3   |                          |           |           |           |           |           |
| 54_4   |                          |           |           |           |           |           |
| 54_5   |                          |           |           |           |           |           |
| 54_6   | 793-802                  | 811-826   | 834-848   | 866-876   | 893-903   | 907-918   |
| 328_1  |                          |           |           |           |           |           |

Table 4

| ORF    | Antigenic Regions (cont) |           |
|--------|--------------------------|-----------|
|        | Region 30                | Region 31 |
| 46_1   |                          |           |
| 63_4   |                          |           |
| 174_6  |                          |           |
| 206_16 |                          |           |
| 267_1  |                          |           |
| 322_1  |                          |           |
| 415_2  |                          |           |
| 214_3  |                          |           |
| 587_3  | 889-911                  | 927-936   |
| 685_1  |                          |           |
| 54_3   |                          |           |
| 54_4   |                          |           |
| 54_5   |                          |           |
| 54_6   | 925-944                  | 951-997   |
| 328_1  |                          |           |

SEQUENCE LISTING

5

(1) GENERAL INFORMATION:

(i) APPLICANT:

10

- (A) NAME: Human Genome Sciences, Inc.
- (B) STREET: 9410 Key West Avenue
- (C) CITY: Rockville
- (D) STATE: Maryland
- (E) COUNTRY: US
- (F) POSTAL CODE: 20850

15

(ii) TITLE OF INVENTION: Staphylococcus aureus Poly-nucleotides and Sequences

20

(iii) NUMBER OF SEQUENCES: 5255

(v) COMPUTER READABLE FORM:

25

- (A) MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
- (B) COMPUTER: HP Vectra 486/33
- (C) OPERATING SYSTEM: MSDOS version 6.2
- (D) SOFTWARE: ASCII Text

30

(vi) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER:
- (B) FILING DATE:

35

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: US 60/009,861
- (B) FILING DATE: 05-JAN-1996

40

(2) INFORMATION FOR SEQ ID NO:1:

45

50

55

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5895 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

|    |  |      |
|----|--|------|
| 10 | TCCATTATGA AGTCACAAGT ACTATAAGCT GCGATGTTAC CAATGTTTTT TAAATCCCA   | 60   |
|    | GTAATAAAAT CAAAAATAA GTTAAATAAT GTATTCATTT TAAGTCCTCC TTAATAAAGa   | 120  |
| 15 | aaataGGTAA TAATGTAATA GCTTCTATTA TGATGCCTAA TTGAATGAAT TGGGCAAATG  | 180  |
|    | GCTCTTTGAT GATAAGTGTG ATAATGAAAA GGGTTAAACT AACATAATC GCATAATATT   | 240  |
|    | TTTTTCGTTT AATAAGTCGC ACAGGAATGG GCTTCTTTTT AGTTGCTGCA GGAGCATATA  | 300  |
| 20 | CTGAGATTAC ACCTAAAGAA ATAAGTGTAA AAATAATCAT AATTAAAAAG TTAATATGAA  | 360  |
|    | AATTTACTAT TACTAAAGGT AAAAGTATAA ATAGTATAAT ACTTTCTACA TAACACCAAA  | 420  |
|    | AAGAAGAAGG TGCATGTGCa CCATGTGCAT GtCTTCTTAT TAAATAAAAT GTTAAATTCG  | 480  |
| 25 | TAATTAACGT AAACAGAAAA ATGTTTAAAA TATAGGCAAT AGTATACATA ACAATTAATT  | 540  |
|    | TACCTATATT TTTAGCTAAG ACCTGCATCC CTAATCGTAC TTGCAAAAAT TGAATATGAT  | 600  |
|    | CTAAGTTATT TCTCTTTTGA AGATACGTGG CAAACTGGTC AATTTTATTA TCAAAATAAT  | 660  |
| 30 | TCAATTTTAC ACCACTCTCC TCACTGTCTAT TATACGATTT AGTACAATCT TTTATCATTa | 720  |
|    | TATTGCCTAA CTGTAGGAAA TAAATACTTA ACTGTTTAAAT GTAATTTGTA TTTAATATTT | 780  |
| 35 | TAACATAAAA AAATTTACAG TTAAGAATAA AAAACGACTA GTTAAGAAAA ATTGGAAAAT  | 840  |
|    | AAATGCTTTT AGCATGTTTT AATATAACTA GATCACAGAG ATGTGATGGA AAATAGTTGA  | 900  |
|    | TGAGFTGTTT AATTTTAAAGA ATTTTATCT TAATTAAGGA AGGAGTGATT TCAATGGCAC  | 960  |
| 40 | AAGATATCAT TTCAACAATC GGTGACTTAG TAAAATGGAT TATCGACACA GTGAACAAAT  | 1020 |
|    | TCACTAAAAA ATAAGATGAA TAATTAATTA CTTTCATTGT AAATTTGTTA TCTTCGTATA  | 1080 |
|    | GTAATAAAAG TATGAGTTAT TAAGCCATCC CAACTTAATA ACCATGTAAA ATTAGCAAGT  | 1140 |
| 45 | GAGTAACATT TGCTAGTAGA GTTAGTTTCC TTGGACTCAG TGCTATGTAT TTTTCTTAAT  | 1200 |
|    | TATCATTACA GATAATTATT TCTAGCATGT AAGCTATCGT AAACAACATC GATTTATCAT  | 1260 |
| 50 | TATTTGATAA ATAAATTTT TTTCATAATT AATAACATCC CCAAAAATAG ATTGAAAAAA   | 1320 |
|    | TAACTGTAAA ACATTCCCTT AATAATAAGT ATGGTCGTGA GCCCCTCCCA AGCTCGCGGC  | 1380 |
| 55 | CTTTTTTGTA ATGAAGAAGG GATGAGTTAA TCATCATTAT GAGACCCGCC GTTAAATAT   | 1440 |

His Lys Ala  
1025

5 (2) INFORMATION FOR SEQ ID NO:5255:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 155 amino acids  
(B) TYPE: amino acid  
10 (C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5255:

20 Gly Glu Lys Cys Met Phe Leu Ala Trp Asn Glu Ile Arg Arg Asn Lys  
1 5 10 15  
Leu Lys Phe Gly Leu Ile Ile Gly Val Leu Thr Met Ile Ser Tyr Leu  
20 25 30  
25 Leu Phe Leu Leu Ser Gly Leu Ala Asn Gly Leu Ile Asn Met Asn Lys  
35 40 45  
Glu Gly Ile Asp Lys Trp Gln Ala Asp Ala Ile Val Leu Asn Lys Asp  
30 50 55 60  
Ala Asn Gln Thr Val Gln Gln Ser Val Phe Asn Lys Lys Asp Ile Glu  
65 70 75 80  
35 Asn Lys Tyr Lys Lys Gln Ala Thr Leu Lys Gln Thr Gly Glu Ile Val  
85 90 95  
Ser Asn Gly His Gln Lys Asp Asn Val Leu Val Phe Gly Val Glu Lys  
100 105 110  
40 Ser Ser Phe Leu Val Pro Ser Leu Ile Glu Gly His Lys Ala Thr Lys  
115 120 125  
Asp Asn Glu Val Leu Ala Asp Glu Thr Leu Lys Asn Lys Gly Leu Lys  
45 130 135 140  
Leu Gly Asp Thr Leu Ser Leu Ser Xaa Xaa Arg  
145 150 155

50

Claims

1. Computer readable medium having recorded thereon a nucleotide sequence of the Staphylococcus aureus genome as depicted in SEQ ID NOS:1-5,191, a representative fragment thereof or a nucleotide sequence at least  
55 95 % identical to a nucleotide sequence depicted in SEQ ID NOS:1-5,191.
2. Computer readable medium having recorded thereon any one of the fragments of SEQ ID NOS:1-5,191 depicted in Tables 2 and 3 or a degenerate variant thereof.

3. The computer readable medium of claim 1, wherein said medium is selected from the group consisting of a floppy disc, a hard disc, random access memory (RAM), read only memory (ROM), and CD-ROM.
- 5 4. The computer readable medium of claim 3, wherein said medium is selected from the group consisting of a floppy disc, a hard disc, random access memory (RAM), read only memory (ROM), and CD-ROM.
5. A computer-based system for identifying fragments of the *Staphylococcus aureus* genome of commercial importance comprising the following elements:
  - 10 (a) a data storage means comprising the nucleotide sequence of SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95% identical to a nucleotide sequence of SEQ ID NOS:1-5,191;
  - (b) search means for comparing a target sequence to the nucleotide sequence of the data storage means of step (a) to identify homologous sequence(s), and
  - 15 (c) retrieval means for obtaining said homologous sequence(s) of step (b).
6. A method for identifying commercially important nucleic acid fragments of the *Staphylococcus aureus* genome comprising the step of comparing a database comprising the nucleotide sequences depicted in SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95% identical to a nucleotide sequence of SEQ ID NOS:1-5,191 with a target sequence to obtain a nucleic acid molecule comprised of a complementary nucleotide sequence to said target sequence, wherein said target sequence is not randomly selected.
- 20 7. A method for identifying an expression modulating fragment of *Staphylococcus aureus* genome comprising the step of comparing a database comprising the nucleotide sequences depicted in SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95% identical to the nucleotide sequence of SEQ ID NOS:1-5,191 with a target sequence to obtain a nucleic acid molecule comprised of a complementary nucleotide sequence to said target sequence, wherein said target sequence comprises sequences known to regulate gene expression.
- 25 8. A protein-encoding nucleic acid fragment of the *Staphylococcus aureus* genome, wherein said fragment comprises the nucleotide sequence of any one of the fragments of SEQ ID NOS:1-5,191 depicted in Tables 2 and 3 or a nucleotide sequence at least 95% identical to such a nucleotide sequence, or a degenerate variant of any of the aforementioned sequences.
- 30 9. The nucleic acid fragment of claim 8 which is DNA.
10. The nucleic acid fragment of claim 8 which is RNA.
11. A vector comprising a fragment of claim 8.
- 40 12. A fragment of the *Staphylococcus aureus* genome, wherein said fragment modulates the expression of an operably linked open reading frame, wherein said fragment consists of the nucleotide sequence from about 10 to 200 bases in length which is 5' to any one of the open reading frames depicted in Tables 2 and 3 or a nucleotide sequence at least 95% identical to such a nucleotide sequence or a degenerate variant of any of the aforementioned sequences.
- 45 13. A vector comprising a fragment of claim 12.
14. A organism which has been altered to contain any one of the fragments of the *Staphylococcus aureus* genome of claim 8.
- 50 15. A method for producing a polypeptide in a host cell comprising the steps of:
  - 55 (a) incubating an organism of claim 14 under conditions where said fragment is expressed to produce said protein, and
  - (b) isolating said protein.
16. An organism which has been altered to contain any one of the fragments of the *Staphylococcus aureus* genome

of claim 12.

17. A method for regulating the expression of a nucleic acid molecule comprising the step of covalently attaching to said nucleic acid molecule a nucleic acid molecule consisting of the nucleotide sequence from about 30 to 300 bases 5' to any one of the fragments of the *Staphylococcus aureus* genome depicted in Seq ID Nos: 1-5, 191 and Tables 2 and 3 or a nucleotide sequence at least 95% identical to such a nucleotide sequence or a degenerate variant of any of the aforementioned sequences.
18. A nucleic acid molecule being a homolog of any of the fragments of the *Staphylococcus aureus* genome of SEQ ID NOS: 1-5, 191 and Tables 2 and 3, wherein said nucleic acid molecule is produced by a process comprising the steps of:
  - (a) screening a genomic DNA library using as a probe a target sequence defined by any of SEQ ID NOS: 1-5, 191 and Tables 2 and 3, including fragments thereof;
  - (b) identifying members of said library which contain sequences that hybridize to said target sequence;
  - (c) isolating the nucleic acid molecules from said members identified in step (b).
19. A DNA molecule being a homolog of any one of the fragments of the *Staphylococcus aureus* genome of SEQ ID NOS: 1-5, 191 and Tables 2 and 3, wherein said nucleic acid molecule is produced by a process comprising the steps of:
  - (a) isolating mRNA, DNA, or cDNA produced from an organism;
  - (b) amplifying nucleic acid molecules whose nucleotide sequence is homologous to amplification primers derived from said fragment of said *Staphylococcus aureus* genome to prime said amplification;
  - (c) isolating said amplified sequences produced in step (b).
20. A polypeptide encoded by a fragment of claim 8.
21. An antibody which selectively binds to any one of the polypeptides of claim 20.
22. A kit for analyzing samples for the presence of polynucleotides derived from *Staphylococcus aureus*, comprising at least one polynucleotide containing a nucleotide sequence of any one of the fragments SEQ ID NOS: 1-5, 191 depicted in Tables 2 and 3 or a nucleotide sequence at least 95% identical thereto or a degenerate variant of any of the aforementioned sequences, that will hybridize to a *staphylococcus aureus* polynucleotide under stringent hybridization conditions, and a suitable container.
23. A *Staphylococcus aureus* polypeptide comprising an amino acid sequence identical to an amino acid sequence selected from the group consisting of SEQ ID NOS: 5, 192 to 5, 255 or comprising an amino acid sequence having at least 95% identity to such a sequence.
24. A *Staphylococcus aureus* polypeptide antigen comprising at least one epitope derived from a *Staphylococcus aureus* polypeptide selected from the group consisting of SEQ ID NOS: 5, 192 to 5, 255.
25. A polypeptide comprising at least one epitope encoded by a *Staphylococcus aureus* amino acid sequence selected from the group consisting of the epitopic sequences listed in Table 4.
26. The polypeptide of claim 24 or 25, wherein said polypeptide is fixed to a solid phase.
27. A diagnostic kit for detecting *Staphylococcus aureus* infection comprising
  - (a) an isolated polypeptide antigen of claim 24, and
  - (b) means for detecting the binding of an antibody contained in a biological fluid to said antigen.
28. A vaccine composition comprising a polypeptide of claim 24 present in a pharmaceutically acceptable carrier.
29. A method of vaccinating an individual against *Staphylococcus aureus* infection comprising, administering to an individual the vaccine composition of claim 28.

Figure 1

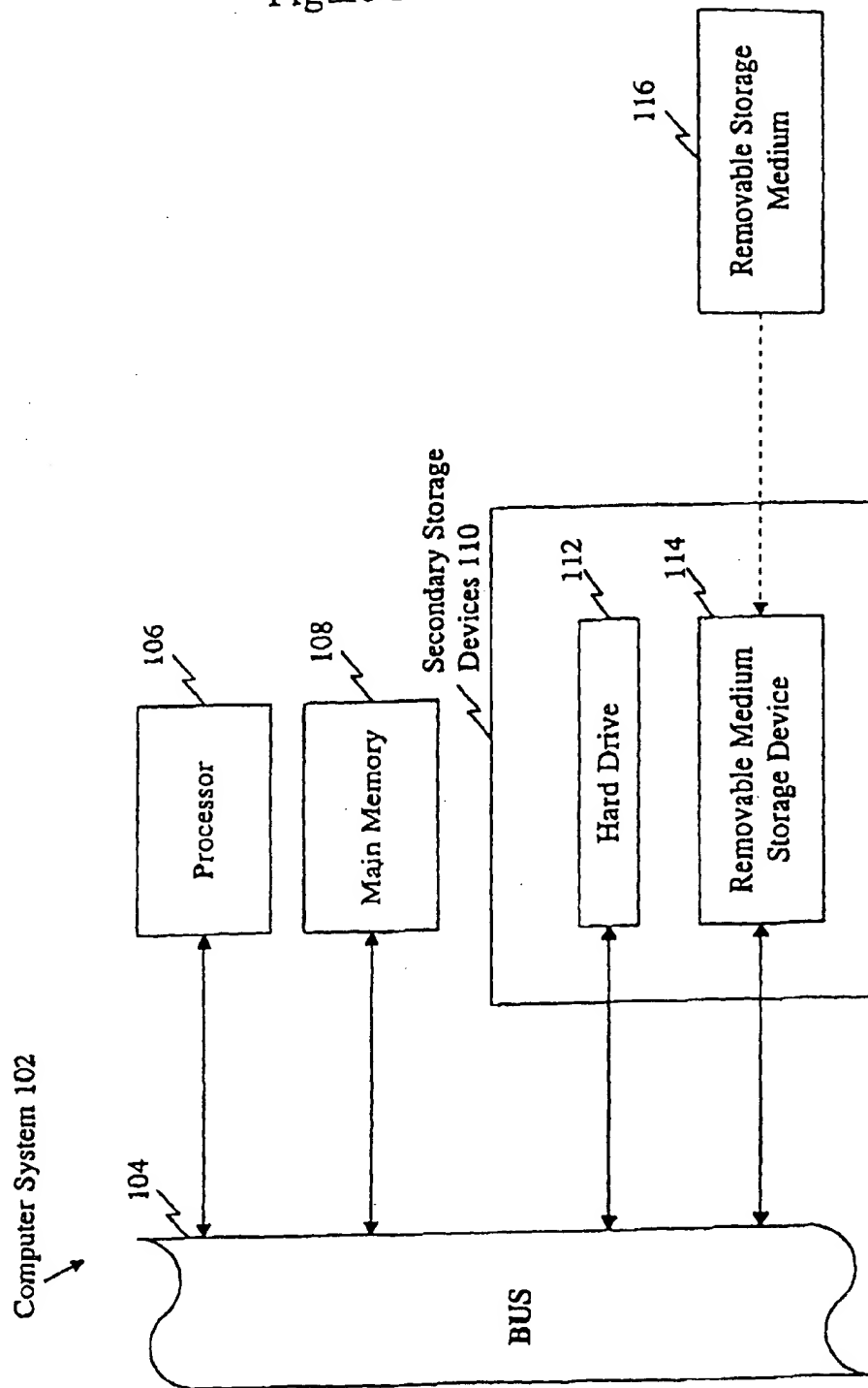


Figure 2

